(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number
WO 01/07471 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/US00/19948

(22) International Filing Date: 21 July 2000 (21.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/145,075	21 July 1999 (21.07.1999)	US
60/153,129	8 September 1999 (08.09.1999)	US
60/164,647	10 November 1999 (10.11.1999)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/145,075 (CIP)
Filed on	21 July 1999 (21.07.1999)
US	60/153,129 (CIP)
Filed on	8 September 1999 (08.09.1999)
US	60/164,647 (CIP)
Filed on	10 November 1999 (10.11.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **AZIMZAI, Yalda** [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **LU, Dyung, Aina, M.** [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **SHAH, Purvi** [IN/US]; 1608 Queen Charlotte Drive #5, Sunnyvale, CA 94087 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/07471 A2

(54) Title: CELL CYCLE AND PROLIFERATION PROTEINS

(57) Abstract: The invention provides human cell cycle and proliferation proteins (CCYPR) and polynucleotides which identify and encode CCYPR. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CCYPR.

CELL CYCLE AND PROLIFERATION PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cell cycle and proliferation proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

BACKGROUND OF THE INVENTION

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms, while in multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Details of the cell division cycle may vary, but the basic process consists of three principal events. The first event, interphase, involves preparations for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions are under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points.

Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of mictrotubules and associated proteins such as dynein, which originate from polar mitotic centers. During metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, mutation studies in the Drosophila melanogaster *zw10* gene show a disruption in chromosome segregation. ZW10 protein appears to function at the kinetochore as a tension-sensing checkpoint during the onset of anaphase. ZW10 appears to have a direct role in the recruitment of dynein to the kinetochore, and, dynein's involvement in the coordination of chromosome separation at the onset of anaphase and/or poleward movement (Starr, D.A. et al. (1998) J. Cell Biol. 142:763-774).

Regulated progression of the cell cycle depends on the integration of growth control pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating pheromone pathway

when they are overexpressed. Known regulators include human CPR (cell cycle progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins including cyclins, tumor suppressor binding proteins, chaperones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M.C. et al. (1997) Genetics 147:1063-1076).

The human CDC protein, CDC23, is homologous to the S. cerevisiae protein CDC23 which functions in the transition from metaphase to anaphase as well as in the exit from mitosis (Zhao, N. et al. (1998) Genomics 53:184-190). The C. elegans gene cullin-1 (cul1) is a negative regulator of the cell cycle. cul1 regulates the G1 to S phase transition and C. elegans cul1 mutants exhibit hyperplasia of all tissues through acceleration of this transition by overriding mitotic arrest. cul1 is a member of a conserved gene family that spans S. cerevisiae, nematodes and humans (Kipreos, E.T. et al. (1996) Cell 85:929-839).

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (Cdks). The Cdks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. There appears to be a single Cdk in Saccharomyces cerevisiae and Schizosaccharomyces pombe whereas mammals have a variety of specialized Cdks. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDC4 and CDC53. In addition, Cdks are further regulated by binding to inhibitors and other proteins such as Suc1 that modify their specificity or accessibility to regulators (Patra, D. and W.G. Dunphy (1996) Genes Dev. 10:1503-1515; and Mathias, N. et al. (1996) Mol. Cell Biol. 16:6634-6643).

Reproduction

The male and female reproductive systems are complex and involve many aspects of growth and development. The anatomy and physiology of the male and female reproductive systems are reviewed in Guyton, A.C. ((1991) Textbook of Medical Physiology, W.B. Saunders Co., Philadelphia PA, pp.899-928).

The male reproductive system includes the process of spermatogenesis, in which the sperm are formed. Male reproductive functions are regulated by various hormones. The hormones exert their effects on accessory sexual organs, and are involved in cellular metabolism, growth, and other bodily functions.

Spermatogenesis begins at puberty as a result of stimulation by gonadotropic hormones released from the anterior pituitary. Immature sperm (spermatogonia) undergo several mitotic cell

divisions before undergoing meiosis and full maturation. The testes secrete several male sex hormones. Testosterone, the most abundant, is essential for growth and division of the immature sperm, and for the masculine characteristics of the male body. Three other male sex hormones, gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH), control sexual function.

The uterus, ovaries, fallopian tubes, vagina, and breasts comprise the female reproductive system. The ovaries and uterus are the source of ova and the location of fetal development, respectively. The fallopian tubes and vagina are accessory organs attached to the top and bottom of the uterus, respectively. Both the uterus and ovaries have additional roles in the development and loss of reproductive capability during a female's lifetime. The primary role of the breasts is lactation. Multiple endocrine signals from the ovaries, uterus, pituitary, hypothalamus, adrenal glands, and other tissues coordinate reproduction and lactation. These signals vary during the monthly menstruation cycle and during the female's lifetime. Similarly, the sensitivity of reproductive organs to these endocrine signals varies during the female's lifetime.

A combination of positive and negative feedback to the ovaries, pituitary and hypothalamus glands controls physiologic changes during the monthly ovulation and endometrial cycles. The anterior pituitary secretes two major gonadotropin hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), regulated by negative feedback of steroids, most notably by ovarian estradiol. If fertilization does not occur, estrogen and progesterone levels decrease. This sudden reduction of the ovarian hormones leads to menstruation, the desquamation of the endometrium.

Hormones further govern all the steps of pregnancy, parturition, lactation, and menopause. During pregnancy large quantities of human chorionic gonadotropin (hCG), estrogens, progesterone, and human chorionic somatomammotropin (hCS) are formed by the placenta. hCG, a glycoprotein similar to luteinizing hormone, stimulates the corpus luteum to continue producing more progesterone and estrogens, rather than to involute as occurs if the ovum is not fertilized. hCS is similar to growth hormone and is crucial for fetal nutrition.

The female breast also matures during pregnancy. Large amounts of estrogen secreted by the placenta trigger growth and branching of the breast milk ductal system while lactation is initiated by the secretion of prolactin by the pituitary gland.

Parturition involves several hormonal changes that increase uterine contractility toward the end of pregnancy, as follows. The levels of estrogens increase more than those of progesterone. Oxytocin is secreted by the neurohypophysis. Concomitantly, uterine sensitivity to oxytocin increases. The fetus itself secretes oxytocin, cortisol (from adrenal glands), and prostaglandins.

Menopause occurs when most of the ovarian follicles have degenerated. The ovary then

produces less estradiol, reducing the negative feedback on the pituitary and hypothalamus glands. Mean levels of circulating FSH and LH increase, even as ovulatory cycles continue. Therefore, the ovary is less responsive to gonadotropins, and there is an increase in the time between menstrual cycles. Consequently, menstrual bleeding ceases, and reproductive capability ends.

Differentiation and Proliferation

Tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

Embryogenesis is a process in which distinct patterns of protein expression control proper development. This process involves a host of proteins each with distinct and highly coordinated expression patterns. For example, in the mouse, temporally regulated expression of two related genes Msg1 and Mrg1 contribute to normal embryonic development. Msg1 is expressed in the posterior domains of the developing mesoderm, while Mrg1 is expressed in the anterior visceral endoderm. Properly coordinated expression of each protein throughout embryogenesis is critical for proper tissue and organ formation (Dunwoodie, S.L. et al. (1998) Mech. Dev. 72:27-40).

Growth factors were originally described as serum factors required to promote cell proliferation. Most growth factors are large, secreted polypeptides that act on cells in their local environment. Growth factors bind to and activate specific cell surface receptors and initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor tyrosine kinases which undergo autophosphorylation upon ligand binding. Autophosphorylation enables the receptor to interact with signal transduction proteins characterized by the presence of SH2 or SH3 domains (Src homology regions 2 or 3) . These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G proteins act as molecular switches that activate other downstream events, such as mitogen-activated protein kinase (MAP kinase) cascades. MAP kinases ultimately activate transcription of mitosis-promoting genes.

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein

coupled receptor (GPCR), found predominantly on the surface of immune, neuronal and neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C, Ca^{2+}, and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects. These signaling cascades often involve activation of kinases and phosphatases. Some growth factors, such as some members of the transforming growth factor beta (TGF-β) family, act on some cells to stimulate cell proliferation and on other cells to inhibit it. Growth factors may also stimulate a cell at one concentration and inhibit the same cell at another concentration. Most growth factors also have a multitude of other actions besides the regulation of cell growth and division: they can control the proliferation, survival, differentiation, migration, or function of cells depending on the circumstance. For example, the tumor necrosis factor/nerve growth factor (TNF/NGF) family can activate or inhibit cell death, as well as regulate proliferation and differentiation. The cell response depends on the type of cell, its stage of differentiation and transformation status, which surface receptors are stimulated, and the types of stimuli acting on the cell (Smith, A. et al. (1994) Cell 76:959-962; and Nocentini, G. et al. (1997) Proc. Natl. Acad. Sci. USA 94:6216-6221).

Neighboring cells in a tissue compete for growth factors, and when provided with "unlimited" quantities in a perfused system will grow to even higher cell densities before reaching density-dependent inhibition of cell division. Cells often demonstrate an anchorage dependence of cell division as well. This anchorage dependence may be associated with the formation of focal contacts linking the cytoskeleton with the extracellular matrix (ECM). The expression of ECM components can be stimulated by growth factors. For example, TGF-β stimulates fibroblasts to produce a variety of ECM proteins, including fibronectin, collagen, and tenascin (Pearson, C.A. et al. (1988) EMBO J. 7:2977-2981). In fact, for some cell types, specific ECM molecules, such as laminin or fibronectin, may act as growth factors. Tenascin-C and -R, expressed in developing and lesioned neural tissue, provide stimulatory/anti-adhesive or inhibitory properties, respectively, for axonal growth (Faissner, A. (1997) Cell Tissue Res. 290:331-341).

Cancers and immune disorders are characterized by uncoordinated cell proliferation. Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which convert normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein, and other oncoproteins are abnormally expressed with respect to location or amount of expression. The latter category of oncoprotein causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal

transducers, nuclear transcription factors, and cell-cycle control proteins. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation (Nigg, E.A. (1995) BioEssays 17:471-480).

Many oncogenes have been identified and characterized. These include sis, erbA, erbB, her-2, mutated G_{α}, src, abl, ras, crk, jun, fos, myc, and mutated tumor-suppressor genes such as RB, p53, mdm2, Cip1, p16, and cyclin D. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22.

Mutations which hyperactivate oncogenes result in cell proliferation. Stimulation of a cell by growth factors activates two sets of gene products, the early-response genes and the delayed-response genes. Early-response gene products include *myc*, *fos*, and *jun*, all of which encode gene regulatory proteins. These regulatory proteins lead to the transcriptional activation of a second set of genes, the delayed-response genes, which include the cell-cycle regulators Cdk and cyclins. For example, the human T-cell leukemia virus type I (HTLV-1) Tax transactivator protein acts as an early response gene by enhancing the activity of a cellular transcription factor. The oncogenic properties of the Tax protein include transformation of primary T-lymphocytes and fibroblasts through cooperation with the a GTP-binding protein, Ras.. Recently investigators have shown that Tax interacts with several PDZ-containing proteins. The PDZ domain, originally described in the Drosophila tumor suppressor protein Discs-Large, is common to membrane proteins thought to be involved in clustering receptors in growth factor signal transduction pathways (Rousset, R. et al. (1998) Oncogene 16:643-654).

Tumor-suppressor genes are involved in regulating cell proliferation. Mutations which cause reduced or loss of function in tumor-suppressor genes result in uncontrolled cell proliferation. For example, the retinoblastoma gene product (RB), in a non-phosphorylated state, binds several early-response genes and suppresses their transcription, thus blocking cell division. Phosphorylation of RB causes it to dissociate from the genes, releasing the suppression, and allowing cell division to proceed.

Other gene products involved in cell proliferation, differentiation, and apoptosis are yet to be discovered. One method currently being utilized to help identify such new molecules involves comparisons between quiescent and proliferative tissues. For example, a subtractive hybridization screen of human placental cytotrophoblast cells identified 20 genes whose expression levels rose due to

EGF induction of cell proliferation. (Morrish, D.W. et al. (1996) Placenta 17:431-441). Another method involves identification of molecules produced in cells treated with anti-tumorigenic agents, such as dithiolethiones. Presumably, the protective action of these anti-tumorigenic agents is associated with the induction of tumor suppressor gene products (Primiano, T. et al. (1996) Carcinogenesis 17:2297-2303).

In another example, the candidate tumor-suppressor gene ING1, that codes a nuclear protein, p33ING1, is involved in the negative regulation of cell proliferation. The action of p33ING1 is dependent upon the activity of another tumor-suppressor gene, p53. p53 is a cellular stress-responsive gene requiring the activity of p33ING1 to effectively induce growth inhibition of cells. p33ING1 and p53 have been shown to physically associate through immunoprecipitation studies (Garkavtsev, I. et al. (1998) Nature 391:295-298).

Apoptosis

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

Aging and Senescence

Studies of the aging process or senescence have shown a number of characteristic cellular and

molecular changes (Fauci, A.S. et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p.37). These characteristics include increases in chromosome structural abnormalities, DNA cross-linking, incidence of single-stranded breaks in DNA, losses in DNA methylation, and degradation of telomere regions. In addition to these DNA changes, post-translational alterations of proteins increase including deamidation, oxidation, cross-linking, and nonenzymatic glycosylation. Still further molecular changes occur in the mitochondria of aging cells through deterioration of structure. These changes eventually contribute to decreased function in every organ of the body.

The discovery of new cell cycle and proliferation proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, cell cycle and proliferation proteins, referred to collectively as "CCYPR" and individually as "CCYPR-1," "CCYPR-2," "CCYPR-3," "CCYPR-4," "CCYPR-5," "CCYPR-6," "CCYPR-7," "CCYPR-8," "CCYPR-9," "CCYPR-10," "CCYPR-11," "CCYPR-12," "CCYPR-13," "CCYPR-14," "CCYPR-15," "CCYPR-16," "CCYPR-17," "CCYPR-18," "CCYPR-19," "CCYPR-20," "CCYPR-21," "CCYPR-22," "CCYPR-23," "CCYPR-24," "CCYPR-25," "CCYPR-26," "CCYPR-27," "CCYPR-28," "CCYPR-29," "CCYPR-30," "CCYPR-31," "CCYPR-32," "CCYPR-33," "CCYPR-34," "CCYPR-35," "CCYPR-36," "CCYPR-37," "CCYPR-38," "CCYPR-39," "CCYPR-40," "CCYPR-41," "CCYPR-42," "CCYPR-43," "CCYPR-44," "CCYPR-45," "CCYPR-46," "CCYPR-47," "CCYPR-48," "CCYPR-49," "CCYPR-50," "CCYPR-51," "CCYPR-52," "CCYPR-53," "CCYPR-54." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-54.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-

54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-54. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:55-108.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of

SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid

sequence selected from the group consisting of SEQ ID NO:1-54, and a pharmaceutically acceptable excipient. In one embodiment, the pharmaceutical composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CCYPR, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group

consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:55-108, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:55-108, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of the above polynucleotide sequence; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CCYPR.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of CCYPR.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding CCYPR were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CCYPR" refers to the amino acid sequences of substantially purified CCYPR obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CCYPR. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CCYPR either by directly interacting with CCYPR or by acting on components of the biological pathway in which CCYPR participates.

An "allelic variant" is an alternative form of the gene encoding CCYPR. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CCYPR include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CCYPR or a polypeptide with at least one functional characteristic of CCYPR. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CCYPR, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CCYPR. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CCYPR. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CCYPR is retained. For example, negatively charged amino

acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CCYPR. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CCYPR either by directly interacting with CCYPR or by acting on components of the biological pathway in which CCYPR participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CCYPR polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as

phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CCYPR, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CCYPR or fragments of CCYPR may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino

acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr.

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of CCYPR or the polynucleotide encoding CCYPR which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:55-108 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:55-108, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:55-108 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:55-108 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:55-108 and the region of SEQ ID NO:55-108 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-54 is encoded by a fragment of SEQ ID NO:55-108. A fragment of SEQ ID NO:1-54 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-54. For example, a fragment of SEQ ID NO:1-54 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-54. The precise length of a fragment of SEQ ID NO:1-54 and the region of SEQ ID NO:1-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular

biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain

DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency

conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CCYPR which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CCYPR which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CCYPR. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CCYPR.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which

comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CCYPR may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CCYPR.

"Probe" refers to nucleic acid sequences encoding CCYPR, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection

programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid,

amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CCYPR, or fragments thereof, or CCYPR itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type

of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at

least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human cell cycle and proliferation proteins (CCYPR), the polynucleotides encoding CCYPR, and the use of these compositions for the diagnosis, treatment, or prevention of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding CCYPR. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each CCYPR were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each CCYPR and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis along with relevant citations, all of which are expressly incorporated by reference herein in their entirety; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding CCYPR. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:55-108 and to distinguish between SEQ ID NO:55-108 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue

categories which express CCYPR as a fraction of total tissues expressing CCYPR. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing CCYPR as a fraction of total tissues expressing CCYPR. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:66 in inflammatory tissues. It should be noted that SEQ ID NO:76 was found to be expressed predominantly in nervous tissue.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding CCYPR were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:61 maps to chromosome 5 within the interval from 141.40 to 142.60 centiMorgans. This interval also contains gene(s) and/or EST(s) associated with corneal dystrophy and deafness.

SEQ ID NO:73 maps to chromosome 2 within the interval from 73.80 to 83.50 centiMorgans. This interval also contains gene(s) and/or EST(s) associated with hereditary nonpolyposis colorectal carcinoma and Muir-Torre syndrome. SEQ ID NO:74 maps to chromosome 19 within the interval from 41.70 to 58.70 centiMorgans. SEQ ID NO:75 maps to chromosome 17 within the interval from 62.90 to 64.20 centiMorgans. This interval also contains gene(s) and/or EST(s) located within the human breast cancer (BRCA1) gene region. SEQ ID NO:76 maps to chromosome 1 within the interval from 143.30 to 153.90 centiMorgans, to chromosome 3 within the interval from 156.20 to 160.00 centiMorgans, and to chromosome X within the interval from 112.80 to 139.40 centiMorgans. The interval on chromosome X from 112.80 to 139.40 centiMorgans also contains gene(s) and/or EST(s) associated with X-linked agammaglobulinaemia.

SEQ ID NO:77 maps to chromosome 23 within the interval from 173.60 to 179.80 centiMorgans, and to chromosome 11 within the interval from 136.90 centiMorgans to q-terminus. SEQ ID NO:78 maps to chromosome 3 within the interval from 200.00 to 213.70 centiMorgans. SEQ ID NO:81 maps to chromosome 7 within the interval from 167.60 centiMorgans to q-terminus. SEQ ID NO:90 maps to chromosome 2 within the interval from 236.10 to 240.20 centiMorgans, to chromosome 3 within the interval from 16.50 to 43.00 centiMorgans, and to chromosome 6 within the interval from 124.20 to 126.50 centiMorgans. SEQ ID NO:91 maps to chromosome 2 within the interval from 22.40 to 40.70 centiMorgans. SEQ ID NO:98 maps to chromosome 8 within the interval from 40.30 to 60.00 centiMorgans. SEQ ID NO:100 maps to chromosome 14 within the interval from 95.50 to 103.70 centiMorgans, and to chromosome 6 within the interval from 158.50 centiMorgans to q-terminus. SEQ ID NO:104 maps to chromosome 18 within the interval from 32.40 to 42.70 centiMorgans. SEQ ID NO:105 maps to chromosome 19 within the interval from 69.90 to 81.20 centiMorgans.

The invention also encompasses CCYPR variants. A preferred CCYPR variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CCYPR amino acid sequence, and which contains at least one functional or structural characteristic of CCYPR.

The invention also encompasses polynucleotides which encode CCYPR. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:55-108, which encodes CCYPR. The polynucleotide sequences of SEQ ID NO:55-108, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CCYPR. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CCYPR. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:55-108 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:55-108. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CCYPR.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CCYPR, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CCYPR, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CCYPR and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CCYPR under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CCYPR or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CCYPR and its derivatives without altering the encoded amino acid sequences

include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CCYPR and CCYPR derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CCYPR or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:55-108 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CCYPR may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a

known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CCYPR may be cloned in recombinant DNA molecules that direct expression of CCYPR, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CCYPR.

The nucleotide sequences of the present invention can be engineered using methods generally

known in the art in order to alter CCYPR-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CCYPR, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CCYPR may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CCYPR itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of CCYPR, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active CCYPR, the nucleotide sequences encoding CCYPR or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CCYPR. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CCYPR. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CCYPR and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CCYPR and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CCYPR. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu,

N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CCYPR. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CCYPR can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CCYPR into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CCYPR are needed, e.g. for the production of antibodies, vectors which direct high level expression of CCYPR may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CCYPR. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of CCYPR. Transcription of sequences encoding CCYPR may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CCYPR may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CCYPR in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CCYPR in cell lines is preferred. For example, sequences encoding CCYPR can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß

glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CCYPR is inserted within a marker gene sequence, transformed cells containing sequences encoding CCYPR can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CCYPR under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CCYPR and that express CCYPR may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CCYPR using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CCYPR is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CCYPR include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CCYPR, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega

(Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CCYPR may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CCYPR may be designed to contain signal sequences which direct secretion of CCYPR through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CCYPR may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CCYPR protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CCYPR activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CCYPR encoding sequence and the heterologous protein sequence, so that CCYPR may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially

available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CCYPR may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

CCYPR of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CCYPR. At least one and up to a plurality of test compounds may be screened for specific binding to CCYPR. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CCYPR, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CCYPR binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CCYPR, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing CCYPR or cell membrane fractions which contain CCYPR are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CCYPR or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CCYPR, either in solution or affixed to a solid support, and detecting the binding of CCYPR to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CCYPR of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CCYPR. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CCYPR activity, wherein CCYPR is combined with at least one test compound, and the activity of CCYPR in the presence of a test compound is compared with the activity of CCYPR in the absence of the test compound. A change in the activity of CCYPR in the presence of the test compound is

indicative of a compound that modulates the activity of CCYPR. Alternatively, a test compound is combined with an in vitro or cell-free system comprising CCYPR under conditions suitable for CCYPR activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CCYPR may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CCYPR or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CCYPR may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CCYPR can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CCYPR is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CCYPR, e.g., by secreting CCYPR in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CCYPR and cell cycle and proliferation proteins. In addition, the expression of CCYPR is closely associated with inflammation, trauma, cell proliferation and cancer. Therefore, CCYPR appears to play a role in immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased CCYPR expression or activity, it is desirable to decrease the expression or activity of CCYPR. In the treatment of disorders associated with decreased CCYPR expression or activity, it is desirable to increase the expression or activity of CCYPR.

Therefore, in one embodiment, CCYPR or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR. Examples of such disorders include, but are not limited to, an immune disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue disorder (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and disorders of immune cell activation; a cell signaling disorder including

endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, teratogenesis, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and fibrocystic breast disease; and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing CCYPR or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CCYPR in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CCYPR may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those listed above.

In a further embodiment, an antagonist of CCYPR may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CCYPR. Examples of such disorders include, but are not limited to, those immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer, described above. In one aspect, an antibody which specifically binds CCYPR may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CCYPR.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CCYPR may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CCYPR including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CCYPR may be produced using methods which are generally known in the art. In particular, purified CCYPR may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CCYPR. Antibodies to CCYPR may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CCYPR or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to

CCYPR have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CCYPR amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CCYPR may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CCYPR-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CCYPR may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CCYPR and its

specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CCYPR epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CCYPR. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CCYPR-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CCYPR epitopes, represents the average affinity, or avidity, of the antibodies for CCYPR. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CCYPR epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CCYPR-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CCYPR, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CCYPR-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding CCYPR, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CCYPR. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CCYPR. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence

complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CCYPR may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and Somia, N. (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in CCYPR expression or regulation causes disease, the expression of CCYPR from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CCYPR are treated by constructing mammalian expression vectors encoding CCYPR and introducing these vectors by mechanical means into CCYPR-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol.

9:445-450).

Expression vectors that may be effective for the expression of CCYPR include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CCYPR may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CCYPR from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CCYPR expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CCYPR under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining

retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CCYPR to cells which have one or more genetic abnormalities with respect to the expression of CCYPR. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CCYPR to target cells which have one or more genetic abnormalities with respect to the expression of CCYPR. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CCYPR to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of

herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CCYPR to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotech. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CCYPR into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CCYPR-coding RNAs and the synthesis of high levels of CCYPR in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CCYPR into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CCYPR.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CCYPR. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CCYPR. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CCYPR expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CCYPR may be therapeutically useful, and in the treament of disorders associated with decreased CCYPR expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CCYPR may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in

altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CCYPR is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CCYPR are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CCYPR. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition which generally comprises an active ingredient formulated with a pharmaceutically

acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such pharmaceutical compositions may consist of CCYPR, antibodies to CCYPR, and mimetics, agonists, antagonists, or inhibitors of CCYPR.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of pharmaceutical compositions may be prepared for direct intracellular delivery of macromolecules comprising CCYPR or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CCYPR or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CCYPR or fragments thereof, antibodies of CCYPR, and agonists, antagonists or inhibitors of CCYPR, which

ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CCYPR may be used for the diagnosis of disorders characterized by expression of CCYPR, or in assays to monitor patients being treated with CCYPR or agonists, antagonists, or inhibitors of CCYPR. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CCYPR include methods which utilize the antibody and a label to detect CCYPR in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CCYPR, including ELISAs, RIAs, and FACS, are known

in the art and provide a basis for diagnosing altered or abnormal levels of CCYPR expression. Normal or standard values for CCYPR expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to CCYPR under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CCYPR expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CCYPR may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CCYPR may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CCYPR, and to monitor regulation of CCYPR levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CCYPR or closely related molecules may be used to identify nucleic acid sequences which encode CCYPR. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CCYPR, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CCYPR encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:55-108 or from genomic sequences including promoters, enhancers, and introns of the CCYPR gene.

Means for producing specific hybridization probes for DNAs encoding CCYPR include the cloning of polynucleotide sequences encoding CCYPR or CCYPR derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CCYPR may be used for the diagnosis of disorders associated with expression of CCYPR. Examples of such disorders include, but are not limited to, an immune disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome

(AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue disorder (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and disorders of immune cell activation; a cell signaling disorder including endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, perturbations of the menstrual cycle, polycystic

ovarian disease, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, teratogenesis, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and fibrocystic breast disease; and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CCYPR may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CCYPR expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CCYPR may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CCYPR may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CCYPR in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CCYPR, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CCYPR, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified

polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CCYPR may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CCYPR, or a fragment of a polynucleotide complementary to the polynucleotide encoding CCYPR, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CCYPR may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CCYPR are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual .

overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of CCYPR include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for CCYPR, or CCYPR or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present

invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CCYPR to quantify the levels of CCYPR expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CCYPR may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map

data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CCYPR on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CCYPR, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CCYPR and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CCYPR, or fragments thereof, and washed. Bound CCYPR is then detected by methods well known in the art. Purified CCYPR can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CCYPR specifically compete with a test compound for binding CCYPR. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CCYPR.

In additional embodiments, the nucleotide sequences which encode CCYPR may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on

properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/145,075, U.S. Ser. No. 60/153,129, and U.S. Ser. No. 60/164,647, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid

(Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions,

references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:55-108. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related

molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CCYPR occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of CCYPR Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:55-108 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:55-108 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available

from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:61, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:81, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:104, and SEQ ID NO:105 are described in The Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:90, and SEQ ID NO:100, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:90, and SEQ ID NO:100 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VI. Extension of CCYPR Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:55-108 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme

(Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:55-108 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such

extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:55-108 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array

elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37 °C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85 °C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the CCYPR-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CCYPR. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CCYPR. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CCYPR-encoding transcript.

X. Expression of CCYPR

Expression and purification of CCYPR is achieved using bacterial or virus-based expression systems. For expression of CCYPR in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express CCYPR upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CCYPR in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CCYPR by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CCYPR is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CCYPR at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CCYPR obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of CCYPR Activity

An assay for CCYPR activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding CCYPR is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine, a radioactive DNA precursor. Where applicable, varying amounts of CCYPR ligand are added to the transfected cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA and CCYPR activity.

XII. Functional Assays

CCYPR function is assessed by expressing the sequences encoding CCYPR at physiologically

elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CCYPR on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CCYPR and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CCYPR and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of CCYPR Specific Antibodies

CCYPR substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CCYPR amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is

synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CCYPR activity by, for example, binding the peptide or CCYPR to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring CCYPR Using Specific Antibodies

Naturally occurring or recombinant CCYPR is substantially purified by immunoaffinity chromatography using antibodies specific for CCYPR. An immunoaffinity column is constructed by covalently coupling anti-CCYPR antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CCYPR are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CCYPR (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CCYPR binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CCYPR is collected.

XV. Identification of Molecules Which Interact with CCYPR

CCYPR, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CCYPR, washed, and any wells with labeled CCYPR complex are assayed. Data obtained using different concentrations of CCYPR are used to calculate values for the number, affinity, and association of CCYPR with the candidate molecules.

Alternatively, molecules interacting with CCYPR are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CCYPR may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent

No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	55	116462	KIDNNOT01	116462H1 (KIDNNOT01), 116462R1 (KIDNNOT01), 116462X304D1 (KIDNNOT01), 1500439F6 (SINTBST01), 2369977F6 (ADRENOT07)
2	56	1210462	BRSTNOT02	260707H1 (HNT2RAT01), 1210462H1 (BRSTNOT02), 1458882F6 (COLNFET02), 1841248T6 (COLNNOT07), 2378362H1 (ISLTNOT01), 3728643F6 (SMCCNON03)
3	57	1305252	PLACNOT02	794067R6 (OVARNOT03), 871989R1 (LUNGAST01), 1235253F1 (LUNGFET03), 1305252F6 (PLACNOT02), 1305252H1 (PLACNOT02), 1703258T6.comp (DUODNOT02), 2678307H1.comp (OVARTUT07), 3221088H1.comp (COLNNON03), 3647280H1 (ENDINOT01)
4	58	1416289	BRAINOT12	639958R6 (BRSTNOT03), 861752H1 (BRAITUT03), 1416289H1 (BRAINOT12), 1416289X310B1 (BRAINOT12), 1416289X310D2 (BRAINOT12), 1947451R6 (PITUNOT01)
5	59	1558289	SPLNNOT04	1558289H1 (SPLNNOT04), 1852450T6 (LUNGFET03), 2396092F6 (THP1AZT01), 2593267F6 (LUNGNOT22), 2632784F6 (COLNTUT15
6	60	1577739	LNODNOT03	181266R1 (PLACNOB01), 1577739H1 (LNODNOT03), 4180022T6 (SINITUT03), 4597046H1 (COLSTUT01), 4860616H1 (PROSTUT09), 4991290H1 (LIVRTUT11), 5059810H1 (CONDTUT02)
7	61	1752768	LIVRTUT01	256106R1 (HNT2RAT01), 2588141H1 (HNT2RAT01), 1312247F1 (COLNFET02), 1344279T6 (PROSNOT11), 1350089H1 (LATRTUT02), 1440718F6 (THYRNOT03), 1752768F6 (LIVRTUT01), 1752768H1 (LIVRTUT01), 1752768T6 (LIVRTUT01), 2079106F6 (ISLTNOT01), SBYA00612U1
8	62	1887228	BLADTUT07	080294F1 (SYNORAB01), 140055F1 (TLYMNOR01), 285207X42 (EOSIHET02), 516882R6 (MMLR1DT01), 1217892T1 (NEUTGMT01), 1887228H1 (BLADTUT07), 4323029H1 (TLYMUNT01)
9	63	1988468	LUNGAST01	072147R6 (THP1PEB01), 496297H1 (HNT2NOT01), 1362109F6 (LUNGNOT12), 1726095F6 (PROSNOT14), 1726095T6 (PROSNOT14), 1988468H1 (LUNGAST01), 1988468T6 (LUNGAST01), 2232471F6 (PROSNOT16)
10	64	2049176	LIVRFET02	2049176H1 (LIVRFET02), 2049176T6 (LIVRFET02), 2049176X321D1 (LIVRFET02)
11	65	2686765	LUNGNOT23	1502858F6 (BRAITUT07), 1956694X315D1 (CONNNOT01), 2022628X307D1 (CONNNOT01), 2686765F6 (LUNGNOT23), 2686765H1 (LUNGNOT23), 2864555H1 (KIDNNOT20), 2887609F6 (SINJNOT02), 3381980H1 (ESOGNOT04)
12	66	3215187	TESTNOT07	151135R6 (PIBRAGT01), 3215187F6 (TESTNOT07), 3215187H1 (TESTNOT07)
13	67	3500375	PROSTUT13	860585R1 (BRAITUT03), 1318501F1 (BLADNOT04), 1419126F1 (KIDNNOT09), 1483246F6 (CORPNOT02), 2238114T6 (PANCTUT02), 2272329H1 (PROSNON01), 3209746F7 (BLADNOT08), 3403213H1 (ESOGNOT03), 4176619H1 (BRAINOT22), 4614606H1 (BRAYDIT01)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
14	68	5080410	LNODNOT11	1270372X300D1 (BRAINOT09), 3460603H1 (293TF1T01), 5080410H1 (LNODNOT11)
15	69	5218248	BRSTNOT35	1808748X15C1 (PROSTUT12), 1808748X16C1 (PROSTUT12), 3391884H1 (LUNGNOT28)
16	70	058336	MUSCNOT01	058336H1 (MUSCNOT01), 058336T6 (MUSCNOT01), g2206766, g2069225
17	71	1511488	LUNGNOT14	1436265F1 (PANCNOT08), 1511488H1 (LUNGNOT14), 1511488T6 (LUNGNOT14), 1850020F6 (LUNGFET03)
18	72	1638819	UTRSNOT06	1282638T1 (COLNNOT16), 1638819F6 (UTRSNOT06), 1638819H1 (UTRSNOT06), 3597071H1 (FIBPNOT01), SBRA03813D1, SBRA04133D1, SBRA03785D1
19	73	1655123	PROSTUT08	1271351F1 (TESTTUT02), 1353234F1 (LATRTUT02), 1655123H1 (PROSTUT08), 2132186R6 (OVARNOT03), 3296525H1 (TLYJINT01), 3354010H1 (PROSNOT28), 3741838F6 (MENTNOT01), 3741838T6 (MENTNOT01), SXAF03528V1
20	74	2553926	THYMNOT03	403261F1 (TMLR3DT01), 1869739F6 (SKINBIT01), 2197242T6 (SPLNFET02), 2553926H1 (THYMNOT03), 2553956T6 (THYMNOT03), 3935528H1 (PROSTUT09), 5263918F6 (CONDTUT02)
21	75	2800717	PENCNOT01	411179F1 (BRSTNOT01), 415284R1 (BRSTNOT01), 1458971F1 (COLNFET02), 1600810H1 (BLADNOT03), 1622005F6 (BRAITUT13), 2173076F6 (ENDCNOT03), 2520087F6 (BRAITUT21), 2800717H1 (PENCNOT01), 5184583H1 (LUNGTMT03), 5435834H1 (SPLNNOT17), 5872662H1 (COLTDIT04)
22	76	5664154	BRAUNOT01	181534F1 (PLACNOB01), SCHA00262V1
23	77	017900	HUVELPB01	017900H1 (HUVELPB01), 092858F1 (HYPONOB01), 1353543F1 (LATRTUT02), 1353543F6 (LATRTUT02), 1428464F1 (SINTBST01), g1616429
24	78	035102	HUVENOB01	035102H1 (HUVENOB01), 077722R1 (SYNORAB01), 995133H1 (KIDNTUT01), 1356968T6 (LUNGNOT09), 1963135R6 (BRSTNOT04), 2659921F6 (LUNGTUT09), 3110603H1 (BRSTNOT17)
25	79	259983	HNT2RAT01	259131R1 (HNT2RAT01), 259983H1 (HNT2RAT01), 268205R1 (HNT2NOT01), 1305726F1 (PLACNOT02)
26	80	926810	BRAINOT04	926810H1 (BRAINOT04), 3490378T6 (EPIGNOT01), 4774848H1 (BRAQNOT01), SBIA01080D1, SBIA04006D1, SBIA02273D1, SBIA01121D1
27	81	1398816	BRAITUT08	056398F1 (FIBRNOT01), 1252138F2 (LUNGFET03), 1294556T1 (PGANNOT03), 1398816H1 (BRAITUT08), 1545328R1 (PROSTUT04)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
28	82	1496820	PROSNON01	996673H1 (KIDNTUT01), 1496820H1 (PROSNON01), 2368484F6 (ADRENOT07), 3071781X303D1 (UTRSNOR01), 3071781X307B1 (UTRSNOR01), 3071781X316B2 (UTRSNOR01), 3071781X316D3 (UTRSNOR01)
29	83	1514559	PANCTUT01	155768H1 (THP1PLB02), 1229952H1 (BRAITUT01), 1337018X11 (COLNNOT13), 1360361H1 (LUNGNOT12), 1365811H1 (SCORNON02), 1514559F6 (PANCTUT01), 1514559H1 (PANCTUT01)
30	84	1620092	BRAITUT13	1620092F6 (BRAITUT13), 1620092H1 (BRAITUT13), 1832842H1 (BRAINON01), 1843815R6 (COLNNOT08), 1843815T6 (COLNNOT08)
31	85	1678765	STOMFET01	1678765F6 (STOMFET01), 1678765H1 (STOMFET01), 2640786H1 (LUNGTUT08), 3542276F6 (TONSNOT03), 4180591H1 (SINITUT03), 4183383H1 (LIVRDIR01), 4349212H1 (TLYMTXT01), 4718559H1 (BRAIHCT02), 5023762H1 (OVARNON03), 5332272H1 (KIDNNOT34), g1665766
32	86	1708229	PROSNOT16	388493R1 (THYMNOT02), 1503519F1 (BRAITUT07), 1708229H1 (PROSNOT16), 1725267F6 (PROSNOT14), 3089258F6 (HEAONOT03)
33	87	1806454	SINTNOT13	406723H1 (EOSIHET02), 821556R1 (KERANOT02), 1649621F6 (PROSTUT09), 1710552H1 (PROSNOT16), 1806454F6 (SINTNOT13), 1806454H1 (SINTNOT13), 2526283H1 (BRAITUT21), 3869969H1 (BMARNOT03)
34	88	1806850	SINTNOT13	270548H1 (HNT2NOT01), 443885R1 (MPHGNOT03), 1257235F1 (MENITUT03), 1337438H1 (COLNNOT13), 1351820F1 (LATRTUT02), 1544066R1 (PROSTUT04), 1806850F6 (SINTNOT13), 1806850H1 (SINTNOT13), 1984108T6 (LUNGAST01), 2921419H1 (SININOT04), 3109392H1 (BRSTTUT15)
35	89	1851534	LUNGFET03	1851534H1 (LUNGFET03), 2407346R6 (BSTMNON02), 2757389R6 (THP1AZS08), 5513454H1 (BRADDIR01), 5629312H1 (PLACFER01)
36	90	1868749	SKINBIT01	1322048F1 (BLADNOT04), 1398330F1 (BRAITUT08), 1437866F6 (PANCNOT08), 1868749F6 (SKINBIT01), 1868749H1 (SKINBIT01), 2279968R6 (PROSNON01), 2684670H1 (LUNGNOT23), 4632232H1 (GBLADIT02), 4951533H2 (ENDVUNT01), 5077673H1 (LNODNOT11), 5388496H1 (BRAINOT19)
37	91	1980010	LUNGTUT03	127747R1 (TESTNOT01), 357561F1 (PROSNOT01), 357561R1 (PROSNOT01), 918017R1 (BRSTNOT04), 1428117F6 (SINTBST01), 1625080F6 (COLNPOT01), 1720753H1 (BLADNOT06), 1932038F6 (COLNNOT16), 1980010H1 (LUNGTUT03), 3112417F6 (BRSTNOT17), 4174704H1 (SINTNOT21), 4238802H1 (SYNWDIT01), 5499543H1 (BRABDIR01), g4337459

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
38	92	2259032	OVARTUT01	475134H1 (MMLR2DT01), 784284R1 (PROSNOT05), 126412H1 (SYNORAT05), 1418710F1 (KIDNNOT09), 1697570T6 (BLADTUT05), 1874051F6 (LEUKNOT02), 2187960T6 (PROSNOT26), 2259032H1 (OVARTUT01), 2259032R6 (OVARTUT01), 3406237H1 (ESOGNOT03), 3441729H1 (PENCNOT06), 3555764H1 (LUNGNOT31), 3728010H1 (SMCCNON03), 3813639H1 (TONSNOT03), 4031501H1 (BRAINOT23), 4274704H1 (PROSTMT01), 4602450H1 (BRSTNOT07), g3327183
39	93	2359526	LUNGFET05	1667182F6 (BMARNOT03), 2359526H1 (LUNGFET05), 2359526X311D1 (LUNGFET05), 2555305F7 (THYMNOT03), 2654667T6 (THYMNOT04), SCHA00290V1, SCHA00266V1, g1748241
40	94	2456494	ENDANOT01	1860223F6 (PROSNOT18), 2456494H1 (ENDANOT01), 2564671H1 (ADRETUT01), 3618339H1 (EPIPNOT01)
41	95	2668536	ESOGTUT02	1513847H1 (PANCTUT01), 1668943F6 (BMARNOT03), 1668943T6 (BMARNOT03), 1721443F6 (BLADNOT06), 2668536H1 (ESOGTUT02), 3438287H1 (PENCNOT05), SBFA00330F1, SCBA05255V1, SCBA01530V1
42	96	2683225	SINIUCT01	196443R6 (KIDNNOT02), 1243440R6 (LUNGNOT03), 1604540F6 (LUNGNOT15), 2072837H1 (ISLTNOT01), 2683225F6 (SINIUCT01), 2683225H1 (SINIUCT01), 3647874H1 (ENDINOT01), 4029178H1 (BRAINOT23)
43	97	2797839	NPOLNOT01	460779T6 (KERANOT01), 782663H1 (MYOMNOT01), 896898R1 (BRSTNOT05), 1218533H1 (NEUTGMT01), 1312923F6 (BLADTUT02), 2473746F6 (THP1NOT03), 2481564H1 (SMCANOT01), 2797839H1 (NPOLNOT01), 3350118H1 (BRAITUT24), 4184264H1 (BRABDIR01), 4401265H1 (TESTTUT03), 4727770H1 (GBLADIT01), 5080203H1 (LNODNOT11), 5524886H1 (LIVRDIR01)
44	98	2959521	ADRENOT09	046696H1 (CORNNOT01), 087727R6 (LIVRNOT01), 138475H1 (LIVRNOT01), 167505H1 (LIVRNOT01), 647975H1 (CARCTXT02), 781084T1 (MYOMNOT01), 972191R6 (MUSCNOT02), 1309196H1 (COLNFET02), 2641117H1 (LUNGTUT08), 2913953H1 (KIDNTUT15), 2959521H1 (ADRENOT09), 2984654H1 (CARGDIT01), 2985141H1 (CARGDIT01), 3138371H1 (SMCCNOT02), 3386016H1 (ESOGNOT04), 3496187H1 (ADRETUT07), 3614426H1 (EPIPNOT01), 4287819H1 (LIVRDIR01), 5395566H1 (LIVRTUT13), g505101
45	99	3082014	BRAIUNT01	182588H1 (PLACNOB01), 645276R6 (BRSTTUT02), 1497811F1 (SINTBST01), 2051505F6 (LIVRFET02), 3082014H1 (BRAIUNT01), 3464112F6 (293TF2T01), 4603079H1 (BRSTNOT07)
46	100	3520701	LUNGNON03	971201H1 (MUSCNOT02), 1544657R6 (PROSTUT04), 1545570H1 (PROSTUT04), 1671030F6 (BMARNOT03), 1671030T6 (BMARNOT03), 2605263F6 (LUNGTUT07), 3520701H1 (LUNGNON03), 3520701R6 (LUNGNON03)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
47	101	4184320	BRADDIT02	2156956F6 (BRAINOT09), 4184253F6 (BRABDIR01), 4184253T6 (BRABDIR01), 4184320H1 (BRADDIT02), 4252542F6 (BRADDIR01)
48	102	4764233	PLACNOT05	4764233H1 (PLACNOT05), 5634642H1 (PLACFER01), g1148809
49	103	4817352	HELATXT03	426993R6 (BLADNOT01), 426993T6 (BLADNOT01), 488301R6 (HNT2AGT01), 3779640H1 (BRSTNOT27), 4817352H1 (HELATXT03)
50	104	5040573	COLHTUT01	1724126F6 (PROSNOT14), 1859337F6 (PROSNOT18), 2026289R6 (KERANOT02), 2026289T6 (KERANOT02), 2122846T6 (BRSTNOT07), 3225302H1 (ADRETUT07), 3322214H1 (PTHYNOT03), 4587178H1 (BRAQNOT01), 4601227H1 (BRSTNOT07), 4885408H1 (LUNLTMT01), 5040573H1 (COLHTUT01)
51	105	5627029	PLACFER01	967988R1 (BRSTNOT05), 1534642T6 (SPLNNOT04), 1700904F6 (BLADTUT05), 1846971R6 (COLNNOT09), 2112727R6 (BRAITUT03), 2112727T6 (BRAITUT03), 2205225F6 (SPLNFET02), 2828475H1 (TLYMNOT03), 3439165F6 (PENCNOT06), 3604622H1 (LUNGNOT30)
52	106	5678487	293TF2T01	1258787F6 (MENITUT03), 1522008F1 (BLADTUT04), 1597992F6 (BLADNOT03), 2057679H1 (BEPINOT01), 2411504H1 (BSTMNON02), 2467956H1 (THYRNOT08), 2739089F6 (OVARNOT09), 2739089T6 (OVARNOT09), 2740762H1 (BRSTTUT14), 2754616H1 (THP1AZS08), 3254971R6 (OVARTUN01), 3487616H1 (EPIGNOT01), 5678487H1 (293TF2T01)
53	107	5682976	BRAENOT02	350492H1 (LVENNOT01), 825361R1 (PROSNOT06), 879866R1 (THYRNOT02), 1667502F6 (BMARNOT03), 1733323F6 (BRSTTUT08), 1876248T6 (LEUKNOT02), 1963215T6 (BRSTNOT04), 2539188H1 (BONRTUT01), 2896448H1 (KIDNTUT14), 3141553H1 (SMCCNOT02), 3374826F6 (CONNTUT05), 3773427H1 (BRSTNOT25), 3779981H1 (BRSTNOT27), 5682976H1 (BRAENOT02), 5546853H1 (TESTNOC01)
54	108	5992432	FTUBTUT02	645878R6 (BRSTTUT02), 1287660F1 (BRAINOT11), 1287660T6 (BRAINOT11), 1417373F6 (BRAINOT12), 1618868F6 (BRAITUT12), 2269980R6 (UTRSNOT02), 2793117F6 (COLNTUT16), 3246793F6 (BRAINOT19), 3592787H1 (293TF5T01), 5992432H1 (FTUBTUT02), g821012

Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
1	145	T10 S93	N15 N38	Signal peptide: M1-Q33 Protein SH3 domain repeat: L8-R99 GLGF signal transduction-related domain: M1-R99		MOTIFS SPSCAN BLAST_PRODOM BLAST_DOMO
2	340	T39 S190 S268 T307 S88 S102 S165 S226 S230 S234 T337		P120 nuclear proliferating cell antigen: N117-K333 Proliferative cell nucleolar protein P120: E26-G293	Proliferating cell nucleolar antigen P120 (g2649749) A. fulgidus	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
3	418	S246 S415 T142 T156 S292 S349 S369 S64 S247 S298	N190 N191 N203 N288 N306		Candidate tumor suppressor p33ING1 (g2829208) H. sapiens	MOTIFS BLAST_GenBank
4	297	T217 T82 S76 S127 S176 T207 S246 Y189	N74	Germ cell-less protein: E96-N297	Germ cell-less protein (g5814404) Mus musculus	MOTIFS BLIMPS_PFAM BLAST_GenBank
5	184	T34 S103 S5 T136	N76		Differentiation factor MDC-3.13 (g3860093) H. sapiens	MOTIFS BLAST_GenBank
6	173	S109 S24 S59 S66 S141 S142 T152			Posterior end mark-5 (g4107015) C. savignyi	MOTIFS BLAST_GenBank

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
7	591	S582 T71 T208 S217 S339 T475 S493 T536 S45 S105 S153 T208 S305 S336 T578 Y93	N374 N425 N534 N585	Signal peptide M1-L64 TPR domain mitosis control E239-P356 TPR repeat V265-K516	Cell division cycle protein 23 homolog (g5541721) A. thaliana	MOTIFS SPSCAN HMMR_PFAM BLAST_DOMO BLAST_GenBank
8	463	T237 S34 T67 T117 T125 S138 T288 T321 S328 S418 T80 S186 S190 S209 S210 T232 T288 S418 T441 S445 Y416	N208	Formin limb deformity: M1-E335	Lymphocyte specific formin related protein (g4101720) M. musculus	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
9	270		N64 N94 N147		Early embryogenesis MRG1 protein (g2570051) M. musculus	MOTIFS BLAST_GenBank
10	255	S180 T49 T53 S97 S152 T201 S210 S23 S97 T145 T216 S225 S228 T231 S242 Y106 Y240		Polyposis locus TB2 homolog: G15-T117 Polyposis locus protein: V13-T117	Similar to polyposis locus protein 1 (g849238) H. sapiens	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
11	533	S227 S412 S505 S7 S17 S65 T349 S442 T29 S72 S89 S358 S442 T446 S505 Y244		TRE oncogene: R56-I277	TRE oncogene-related protein (g2286196) D. melanogaster	MOTIFS BLOCKS_DOMO BLAST_GenBank

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
12	160	S40		Signal peptide: M1-A30 Transmembrane domain: A6-I29 Cornichon developmental protein: M1-S160	Cornichon-like protein (g4521254) M. musculus	MOTIFS SPSCAN HMMR BLAST_PRODOM BLAST_DOMO BLAST_GenBank
13	531	S195 T196 S357 T45 S172 T199 S212 S322 S465 T495 T45 T241 S255 T279 T319 S328	N244 N401		Cdc 73p (g632679) S. cerevisiae	MOTIFS BLAST_GenBank
14	165	S3 T67 S104			Wolf-Hirschhorn syndrome candidate 2 protein (g3860187) H. sapiens	MOTIFS BLAST_GenBank
15	199	S2 S21 S69 T102 S189			Developmental protein DG1118 (g3789911) D. discoideum	MOTIFS BLAST_GenBank
16	168	S141 S55 S61 T79	N77	Signal peptide M1-S61 H-Rev protein homolog P15-K166	g3777529 retinoic acid receptor responder 3 Homo sapiens	BLAST-GenBank SPSCAN BLAST-PRODOM MOTIFS
17	162	S70 S85 T16 T28 T65 T80 T100 S127 Y111			g207250 growth and transformation dependent protein Rattus norvegicus	BLAST-GenBank

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
18	246	T209 S227 T243 T28 S223 S51 S136 S201	N26 N158	Protein cell intergenic region FTSJ K25-K241	g2622903 cell division protein J Methanobacterium thermoautotrophicum	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS
19	483	T394 T85 S86 S219 S225 T230 S298 T299 T472 S114 S200 T273 S371 T407 T424 T431		Signal peptide M1-G29 OS-9 precursor L54-E281	g1322234 OS-9 precursor Homo sapiens	BLAST-GenBank SPSCAN BLAST-PRODOM MOTIFS
20	280	T129 T6 T102 T119 T181 S250 S46 T72 T84 S262		Signal peptide M1-L28	g3901272 ZW10 interactor Zwint Homo sapiens	BLAST-GenBank SPSCAN MOTIFS
21	425	S122 S235 T60 S192 S203 S204 S218 S226 S307 T313 S332 S366 S370 T375 T402 S409 S89 S118 S241 S284 T360 Y399	N190 N311		g455719 Activated c-raf oncogenic fusion protein homolog Homo sapiens	BLAST-GenBank
22	128	S3 S107	N42	Prenyl group binding site (CAAX box) C125-P128 Ovarian granulosa cell 13.0 KD protein HGR74 N16-P128	g4580592 brain expressed X-linked protein 2 Mus musculus	BLAST-GenBank MOTIFS BLAST-PRODOM
23	113	S88 T20 T37		Biotin-requiring enzyme attachment site: L40-L90	LDOC-1 protein g3869127 (Homo sapiens) Nagasaki,K. et al. (1999) Cancer Lett. 140:227-234.	BLAST-GenBank PROFILESCAN MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
24	308	S95 T79 T98 S184 S246 S251 T55 S184 S226 S294 S300 Y127	N77	Melanoma antigen gene (MAGE) family: M1-Q200, H205-D283, D91-A287	Breast cancer associated gene 1 g4928044 (Homo sapiens) Lurquin,C. et al. (1997) Genomics 46:397-408.	BLAST-GenBank BLAST-PRODOM HMMER-PFAM BLAST-DOMO MOTIFS
25	221	S145 S160 S217 S25 S31 S70 S85 T89 S153 S197 Y34	N139	Annexin VI signature: L86-V95 Sushi domain: T165-C174	Teratocarcinoma expressed gene Tera g1575505 (Mus musculus)	BLAST-GenBank BLIMPS-PRINTS BLIMPS-PFAM MOTIFS
26	402	T344 S39 S78 S109 S237 T269 S273 T376 T381 T383 S11 S49 T89 T344 S364	N76 N107 N171 N362		Paraneoplastic cancer-testis-brain antigen g6179740 (Homo sapiens)	BLAST-GenBank MOTIFS
27	93	S11			Hypoxia inducible gene-1 g4929330 (Homo sapiens)	BLAST-GenBank MOTIFS
28	353	S125 T42 S43 S85 S212 S283 S314 T42 S49 S105 S120 S133 S162 S163 S212 S290	N145 N157 N191	af-4 (FEL protein): S195-K353 E4-Q185	AF5q31 protein g6601438 (Homo sapiens)	BLAST-GenBank BLAST_PRODOM BLAST-DOMO MOTIFS
29	120	T57		Cyclin-dependent kinase inhibitor: D7-P106, M1-N114	Cyclin dependent kinase inhibitor CIP1 g2276312 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
30	144	S15 S64		Transmembrane domain: I93-I110	Transformation dependent protein g207250 (Rattus norvegicus) N.Glaichenhaus and F.Cuzin (1987) Cell 50:1081-1089.	BLAST-GenBank MOTIFS HMMER
31	933	S603 T51 S109 T129 S162 S203 S223 S224 S240 S261 S266 S280 S282 S313 T328 S346 S353 S378 S394 S460 S491 S499 T531 S627 S641 S642 S725 T732 S759 S188 S309 S423 S592 S671 S675 T706 S771 Y856	N107 N238 N639 N883		Replication protein Smp2 g218488 (Saccharomyces cerevisiae) Irie,K. et al. (1993) Mol. Gen. Genet. 6:283-288.	BLAST-GenBank MOTIFS
32	268	S7 T104 T154 S169	N90	Serine-Threonine kinase Binder MPS1: L74-I230	Putative mitotic protein (Schizosaccharomyces pombe) g3947877 F.C.Luca and M.Winey (1998) Mol Biol Cell 9:29-46.	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
33	337	T29 S236 T44 T238		Leucine zipper: L259-L280, L266-L287	DNA binding protein g184390 (Homo sapiens) Weitzel,J.N. et al. (1992) Genomics 14:309-319.	BLAST-GenBank MOTIFS
34	565	T17 S34 S61 S66 T138 T142 S174 T238 S245 S265 S436 S466 S527 S106 S205 S218 S258 T297 S314 T325 S463 T470 Y460	N347 N386 N506	F-Box domain: H75-Y123, L82-N95 Disease resistance protein: G254-I270	F-box protein FLR1 g7672734 (Homo sapiens)	BLAST-GenBank HMMER_PFAM BLIMPS-PRINTS MOTIFS
35	228	S200 T47 T62 S78 S107 S188 S192 S206 S200 S205 S213	N36 N94 N225		Predicted WHSC1 protein (Wolf-Hirschhorn syndrome critical region 1) g4378022 (Homo sapiens) Stecc I. et al. (1998) Hum. Mol. Genet. 7:1071-1082.	BLAST-GenBank MOTIFS
36	495	S451 S152 S365 S478 S108 S171 S181 T192 T347 T409 S435 Y86 Y111 Y203			Malignant brain tumor protein 1(3)mbt g3811111 (Homo sapiens) Koga,H. et al. (1999) Oncogene 18:3799-3809.	BLAST-GenBank MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
37	1336	T635 T769 S902 S10 S32 S33 T76 S95 S156 T298 S313 T427 S467 T579 T626 T642 S661 T668 S680 T699 T729 S774 S834 T859 T915 S944 S959 S961 S997 S1049 T1085 S1132 S1227 T1245 S1249 T48 S94 T169 S224 T352 T379 T389 T475 T696 S867 T883 T889 S940 S961 S1220 Y631	N148 N152 N345 N385 N1213 N1247	Ribosomal protein S14 signature: R1172-N1194 Leucine zipper: L211-L232	Neuroblastoma related protein g4337460 (Homo sapiens)	BLAST-GenBank BLIMPS-PRINTS MOTIFS
38	934	T532 S11 T23 T80 S171 S202 T214 T240 S244 T275 S412 S416 S437 S518 T523 S719 S746 S753 S796 S807 S93 T279 T527 S598 T780	N8 N210 N426	SAP: I92-Q364	Sap2 family putative cell cycle dependent phosphatase g3426127 (Schizosaccharomyces pombe) Luke,M.M. et al. (1996) Mol. Cell Biol. 16:2744-2755.	BLAST-GenBank BLAST-DOMO MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
39	515	T72 S122 S175 S272 S277 S305 T420 S422 T432 T79 S139 T189 S215 T316 S457 T486 Y13 Y383	N16 N31 N115	Metastasis-Associated Protein: E65-R230 Leucine zipper: L234-L255	Metastasis associated gene g1008544 (Homo sapiens) Toh,Y. et al. (1995) Gene 159:97-104 Toh,Y, et al. (1994) J Biol. Chem. 269:22958-22963.	BLAST-GenBank BLAST-PRODOM BLIMPS-PRINTS MOTIFS
40	146	S61		Leucine zipper: L5-L26, L12-L33, L19-L40	LDOC1 g3869127 (Homo sapiens)	BLAST-GenBank BLIMPS-PFAM MOTIFS
41	580	S324 S36 S340 S550 S86 T109 T119 T150 T226 S329 S340	N190	Cyclin: H19-K262	Cyclin K g3746549 (Homo sapiens) Edwards,M.C. et al. (1998) Mol. Cell Biol. 18:4291-4300.	BLAST-GenBank BLAST-PRODOM MOTIFS
42	131	S78 T121 T26		Presenilin: Q64-K75	Cell growth regulator DRR1 g4322559 (Homo sapiens) G.Thomas and M.N.Hall (1997) Curr. Opin. Cell Biol. 9:782-787.	BLAST-GenBank BLIMPS-PRINTS MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
43	812	S44 S588 S646 S801 S111 S120 S134 T140 S148 S150 S181 T185 S262 S279 S440 T477 S497 T520 T542 T605 S675 S40 T64 T311 T316 T319 T505 S562 S565 T566 T695 S702 S707 S708 T739 T776 S790 Y277	N503 N618	NOL1/NOP2/fmu(sun) family signature: F454-G467, F300-K585, I388-M402, G410-G433, F454-G467, K507-L532, E189-M576 Proliferating Cell Nucleolar Antigen P120: M1-S134, E135-T311, F587-G805	Proliferating cell nuclear protein P120 g287723 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DOMO BLIMPS-BLOCKS MOTIFS HMMER-PFAM
44	537	S505 T69 S138 S194 S310 S337 S356 T386 S485 S37 T45 T282	N122 N132 N147	Transmembrane domains: I506-G532, V271-L290, W472-F490	Estrogen induced protein in breast cancer LIV-1 g1256001 (Homo sapiens)	BLAST-GenBank HMMER MOTIFS
45	584	S185 T324 S343 T537 S575 S17 T102 S128 T229 T374 S412 T450	N28	Cytochrome C motif: C283-T288 Metastasis-associated protein MTA1: R19-R143, D144-K321, G340-G483, P432-K555 Leucine zipper: L147-L168	Metastasis associated gene g1008544 (Homo sapiens) Toh,Y. et al. (1995) Gene 159:97-104 Toh,Y. et al. (1994) J. Biol. Chem. 269:22958-22963.	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
46	425	S190 T301 S12 S19 S41 S205 T206 T235 S263 S265 T315 S43 S52 S85 T93 T351 S411 Y422	N275	MLO2 mitosis-associated protein: L24-R188, P226-Y245, N308-E408		BLAST-PRODOM MOTIFS
47	255	T9 T147 S237	N144	Melastatin: M1-R172, G199-G255	Melastatin g3047242 (Mus musculus) Duncan,L.M. et al. (1998) Cancer Res. 58:1515-1520.	BLAST-GenBank BLAST-PRODOM MOTIFS
48	111	T30 S2 T8			Melanoma associated antigen GAGE-8 g3511023 (Homo sapiens) Van den Eynde,B. et al. (1995) J. Exp. Med. 182:689-698.	BLAST-GenBank MOTIFS
49	422	T110 T159 S136 S150 T163 T190 S383 T413 S9 T27 S46 S96 T347 S359 S363 S368 Y350		XPMC2 (mitosis associated inducing protein): A236-E402	Mitotic regulator XPMC2 (Xenopus gene which prevents mitotic catastrophe) g595380 (Xenopus laevis) J.Y.Su and J.L.Maller (1995) Mol. Gen. Genet. 246:387-396.	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
50	397	S20 S21 T395 T57 S59 T64 S127 S208 T210 S262 S307 T341 T64 T168 S180 S185 S218 S231 S288 S326	N222 N260	Transmembrane motifs: I361-L380, L24-L44 Cell division control protein: K17-L347	Cell cycle protein CDC1 g550426 (Saccharomyces cerevisdiae)	BLAST-GenBank HMMER BLAST-PRODOM MOTIFS
51	800	S56 S448 T721 S760 S48 S84 S111 S119 T146 T189 T235 S250 S265 T275 S321 S335 T392 S448 T466 S474 T562 S596 S598 T626 S686 S3 S4 S65 S89 S107 T123 S348 T398 T402 T716 S730 S738 T743 S789 Y102 Y316 Y569 Y685	N554 N665	Signal peptide: M1-A25 Leucine zipper: L365-L386	SART-1 g4126469 (Mus musculus)	BLAST-GenBank SPSCAN MOTIFS
52	713	S100 T631 S8 T9 S20 T42 T114 T121 T172 T177 T191 T192 S218 T231 T256 S325 S335 S381 T464 T482 T538 T581 T617 S693 S94 S166 T201 S202 S321 T568 S614 T658 Y459	N7 N49 N462	Leucine zipper: L680-L701	Colon cancer antigen NY-CO-8 g3170180 (Homo sapiens) Scanlan, M.J. et al. (1998) Int. J. Cancer 76:652-658.	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
53	880	S18 S68 T123 T143 S159 T178 T286 S294 S327 S376 S388 T397 T403 S426 S438 S474 S563 T587 T634 T645 S659 S665 S677 S756 S799 S809 T827 S870 S82 T88 S99 T131 T165 S215 S253 S362 S487 T510 S525 S589 T593 S622	N60 N251 N338 N514 N585 N643	MybI DNA-binding domain: W808-I816 WD40 domains: L41-N79, K84-N124, T131-D170, G239-D281, A771-S809, F157-T171 Acidic Serine Cluster Repeat: A423-R697	homologous to mouse gene PC326 g458692 (Homo sapiens) Bergsagel, P.L. et al. (1992) Oncogene 7:2059-2064.	BLAST-GenBank BLAST-DOMO HMMER-PFAM BLIMPS-PRINTS MOTIFS
54	855	T460 S8 S179 S261 T288 T313 T377 T706 T719 T755 S764 S803 S851 S34 S67 T129 S190 S339 T391 S483 S502 S537 Y92	N552	Crooked neck protein (RNA processing associated, contains TPR repeat): W398-V814	Predicted TPR domain protein G2315362 (Caenorhabditis elegans) Zhang,K. et al. (1991) Genes Dev. 5:1080-1091.	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 3

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
55	263-307	Cardiovascular (0.200) Gastrointestinal (0.200) Reproductive (0.200)	Cancer (0.433) Inflammation (0.267) Cell Proliferation (0.200)	PBLUESCRIPT
56	406-450	Reproductive (0.222) Cardiovascular (0.167) Gastrointestinal (0.167) Nervous (0.167)	Cancer (0.500) Inflammation (0.389) Cell Proliferation (0.167)	PSPORT1
57	1001-1045	Reproductive (0.265) Gastrointestinal (0.206) Nervous (0.206)	Cancer (0.412) Inflammation (0.324) Cell Proliferation (0.176)	pINCY
58	226-270	Nervous (0.316) Hematopoietic/Immune (0.211) Reproductive (0.211)	Cancer (0.368) Inflammation (0.368) Cell Proliferation (0.158)	pINCY
59	406-450	Hematopoietic/Immune (0.500) Cardiovascular (0.227)	Cancer (0.182) Inflammation (0.682) Cell Proliferation (0.136)	pINCY
60	56-100	Gastrointestinal (0.545) Nervous (0.182) Reproductive (0.182)	Cancer (0.545) Inflammation (0.364) Cell Proliferation (0.273)	pINCY
61	1046-1090	Nervous (0.271) Reproductive (0.220) Gastrointestinal (0.153)	Cancer (0.542) Inflammation (0.288) Cell Proliferation (0.220)	pINCY
62	226-270	Hematopoietic/Immune (0.288) Nervous (0.178) Reproductive (0.164)	Cancer (0.397) Inflammation (0.548)	pINCY
63	559-603	Reproductive (0.260) Gastrointestinal (0.145) Cardiovascular (0.130)	Cancer (0.458) Inflammation (0.359) Cell Proliferation (0.176)	PSPORT1
64	12-56	Reproductive (0.385) Gastrointestinal (0.231) Cardiovascular (0.154) Nervous (0.154)	Cancer (0.538) Inflammation (0.154) Cell Proliferation (0.154)	pINCY
65	488-532 1091-1135	Reproductive (0.308) Nervous (0.282) Gastrointestinal (0.154)	Cancer (0.487) Inflammation (0.231) Cell Proliferation (0.103)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
66	37-81	Nervous (0.500) Dermatologic (0.250) Reproductive (0.250)	Inflammation (0.500)	pINCY
67	326-370 1136-1180	Nervous (0.237) Reproductive (0.237) Hematopoietic/Immune (0.158)	Cancer (0.395) Inflammation (0.316) Cell Proliferation (0.158)	pINCY
68	451-495	Nervous (0.312) Reproductive (0.312) Developmental (0.125) Hematopoietic/Immune (0.125) Urologic (0.125)	Cancer (0.562) Inflammation (0.188) Cell Proliferation (0.312)	pINCY
69	64-108	Reproductive (0.233) Nervous (0.174) Cardiovascular (0.140)	Cancer (0.477) Inflammation (0.279) Cell Proliferation (0.198)	pINCY
70	77-121	Cardiovascular (0.500) Musculoskeletal (0.500)	Cancer (0.500) Trauma (0.500)	PBLUESCRIPT
71	164-208	Developmental (0.222) Nervous (0.222)	Cancer (0.444) Cell proliferation (0.222) Trauma (0.222)	pINCY
72	604-648	Reproductive (0.362) Gastrointestinal (0.149) Hematopoietic/Immune (0.128)	Cancer (0.426) Inflammation/Trauma (0.276) Cell proliferation (0.170)	pINCY
73	106-150 1066-1110	Reproductive (0.307) Nervous (0.202) Cardiovascular (0.114)	Cancer (0.482) Inflammation/Trauma (0.307) Cell proliferation (0.175)	pINCY
74	651-695	Hematopoietic/Immune (0.290) Reproductive (0.226) Cardiovascular (0.161)	Inflammation/Trauma (0.451) Cell proliferation (0.230) Cancer (0.320)	pINCY
75	241-285 535-579	Reproductive (0.193) Cardiovascular (0.169) Gastrointestinal (0.157)	Cancer (0.458) Inflammation/Trauma (0.337) Cell proliferation (0.169)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
76	173-217 593-637	Nervous (0.513) Reproductive (0.167)	Inflammation/Trauma (0.371) Cancer (0.333) Cell proliferation (0.141)	pINCY
77	13-57	Reproductive (0.241) Nervous (0.202) Cardiovascular (0.140)	Cancer (0.461) Inflammation (0.180) Cell Proliferation (0.167)	PBLUESCRIPT
78	176-220	Nervous (0.279) Reproductive (0.235) Gastrointestinal (0.147)	Cancer (0.500) Inflammation (0.176) Cell Proliferation (0.162)	PBLUESCRIPT
79	79-123	Nervous (0.280) Cardiovascular (0.160) Developmental (0.160)	Cancer (0.480) Cell Proliferation (0.480) Inflammation (0.160)	PBLUESCRIPT
80	870-914	Nervous (0.571) Reproductive (0.238) Developmental (0.095)	Cancer (0.238) Inflammation (0.381) Cell Proliferation (0.190)	PSPORT1
81	149-194	Nervous (0.216) Reproductive (0.201) Gastrointestinal (0.185)	Cancer (0.432) Inflammation (0.259) Cell Proliferation (0.154)	pINCY
82	150-194	Reproductive (0.375) Cardiovascular (0.125) Endocrine (0.125) Hematopoietic/Immune (0.125) Developmental (0.125) Urologic (0.125)	Cancer (0.375) Inflammation (0.375) Trauma (0.250)	PSPORT1
83	177-221	Reproductive (0.199) Gastrointestinal (0.173) Hematopoietic/Immune (0.128) Nervous (0.128)	Cancer (0.429) Inflammation (0.270) Cell Proliferation (0.186)	pINCY
84	342-386	Reproductive (0.252) Gastrointestinal (0.196) Nervous (0.161)	Cancer (0.483) Inflammation (0.238) Cell Proliferation (0.161)	pINCY
85	124-168	Hematopoietic/Immune (0.308) Cardiovascular (0.154) Nervous (0.154) Gastrointestinal (0.154)	Cancer (0.538) Inflammation (0.308)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
86	238-282	Reproductive (0.277) Cardiovascular (0.181) Nervous (0.169)	Cancer (0.434) Inflammation (0.193) Cell Proliferation (0.157)	PINCY
87	117-161	Reproductive (0.250) Gastrointestinal (0.250) Hematopoietic/Immune (0.115)	Cancer (0.558) Inflammation (0.192) Cell Proliferation (0.115) Trauma (0.115)	pINCY
88	139-183	Nervous (0.237) Reproductive (0.214) Gastrointestinal (0.168)	Cancer (0.397) Inflammation (0.298) Trauma (0.137)	pINCY
89	184-228 352-396	Reproductive (0.556) Nervous (0.222) Hematopoietic/Immune (0.111) Developmental (0.111)	Cancer (0.444) Inflammation (0.333) Cell Proliferation(0.333)	pINCY
90	69-113 879-923	Nervous (0.316) Reproductive (0.193) Hematopoietic/Immune (0.158)	Cancer (0.439) Inflammation (0.211) Cell Proliferation(0.123)	pINCY
91	72-116	Nervous (0.211) Reproductive (0.197) Gastrointestinal (0.158)	Cancer (0.461) Inflammation (0.263) Cell Proliferation(0.211)	PSPORT1
92	489-533	Reproductive (0.274) Nervous (0.217) Gastrointestinal (0.123)	Cancer (0.481) Inflammation (0.189) Cell Proliferation(0.160)	PSPORT1
93	761-805	Reproductive (0.219) Hematopoietic/Immune (0.156) Developmental (0.125)	Cancer (0.312) Cell Proliferation(0.281) Inflammation (0.188) Trauma (0.188)	PSPORT1
94	126-170	Reproductive (0.379) Nervous (0.241) Developmental (0.138)	Cancer (0.414) Cell Proliferation(0.241) Inflammation (0.103)	PBLUESCRIPT
95	1173-1217	Reproductive (0.192) Gastrointestinal (0.192) Nervous (0.173)	Cancer (0.481) Inflammation (0.250) Cell Proliferation(0.212)	pINCY
96	465-509	Hematopoietic/Immune (0.250) Cardiovascular (0.158) Gastrointestinal (0.145)	Inflammation (0.368) Cancer (0.355) Cell Proliferation(0.132)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
97	2427-2471	Nervous (0.224) Reproductive (0.197) Gastrointestinal (0.184)	Cancer (0.474) Cell Proliferation(0.263) Inflammation (0.237)	pINCY
98	23-67	Gastrointestinal (0.270) Reproductive (0.190) Cardiovascular (0.135)	Cancer (0.429) Inflammation (0.278) Cell Proliferation(0.143)	pINCY
99	106-150	Gastrointestinal (0.263) Reproductive (0.263) Nervous (0.158)	Cancer (0.474) Inflammation (0.368) Cell Proliferation(0.211)	pINCY
100	73-117 460-504	Hematopoietic/Immune (0.211) Reproductive (0.211) Cardiovascular (0.105) Developmental (0.105) Gastrointestinal (0.105) Musculoskeletal (0.105)	Cancer (0.474) Inflammation (0.263) Cell Proliferation(0.211)	PSPORT1
101	861-905	Developmental (0. 333) Nervous (0.667)	Cell Proliferation(0. 333) Trauma (0. 333) Neurological (0.333)	pINCY
102	8-52	Developmental (1.000)	Cell Proliferation (1.000)	pINCY
103	199-243	Hematopoietic/Immune (0.143) Nervous (0.179) Reproductive (0.286)	Cancer (0.536) Inflammation (0.250) Cell Proliferation(0.214)	pINCY
104	413-457 908-952	Nervous (0.236) Reproductive (0.222) Gastrointestinal (0.125)	Cancer (0.458) Inflammation (0.236) Cell Proliferation(0.139)	pINCY
105		Reproductive (0.270) Gastrointestinal (0.169) Hematopoietic/Immune 0.101) Developmental (0.101) Nervous (0.101)	Cancer (0.449) Inflammation (0.281) Cell Proliferation(0.258)	pINCY
106	255-299 513-557	Reproductive (0.216) Gastrointestinal (0.196) Nervous (0.157)	Cancer (0.490) Inflammation (0.176) Cell Proliferation(0.176)	pINCY
107	167-211 814-859 1922-1966	Reproductive (0.263) Nervous (0.162) Gastrointestinal (0.141)	Cancer (0.455) Inflammation (0.202) Trauma (0.131)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
108	877-921 2230-2274	Reproductive (0.299) Nervous (0.206) Gastrointestinal (0.134)	Cancer (0.536) Inflammation (0.227) Cell Proliferation(0.124)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
55	KIDNNOT01	Library was constructed using RNA isolated from the kidney tissue of a 64-year-old Caucasian female, who died from an intracranial bleed. Patient history included rheumatoid arthritis.
56	BRSTNOT02	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocystic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.
57	PLACNOT02	Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).
58	BRAINOT12	Library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
59	SPLNNOT04	Library was constructed using RNA isolated from the spleen tissue of a 2-year-old Hispanic male, who died from cerebral anoxia.
60	LNODNOT03	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.
61	LIVRTUT01	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
62	BLADTUT07	Library was constructed using RNA isolated from bladder tumor tissue removed from the anterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated a grade 3 transitional cell carcinoma in the left lateral bladder. Patient history included angina, emphysema, and tobacco use. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
63	LUNGAST01	Library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.
64	LIVRFET02	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
65	LUNGNOT23	Library was constructed using RNA isolated from left lobe lung tissue removed from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia.
66	TESTNOT07	Library was constructed using RNA isolated from testicular tissue removed from a 31-year-old Caucasian male during an unilateral orchiectomy (excision of testis). Pathology indicated a mass containing a large subcapsular hematoma with laceration of the tunica albuginea. The surrounding testicular parenchyma was extensively necrotic.
67	PROSTUT13	Library was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3). Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis,. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis
68	LNODNOT11	Library was constructed using RNA isolated from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Patient history included bronchitis.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
69	BRSTNOT35	Library was constructed using RNA isolated from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated normal breast parenchyma, bilaterally. The patient presented with hypertrophy of breast and headache. Patient history included obesity,lumbago, glaucoma, and alcohol abuse. Family history included cataract,osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, alcoholiccirrhosis of the liver, cerebrovascular disease, and type II diabetes.
70	MUSCNOT01	Library was constructed at Stratagene (STR937209), using RNA isolated from the skeletal muscle tissue of a patient with malignant hyperthermia.
71	LUNGNOT14	Library was constructed using RNA isolated from lung tissue removed from the left lower lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a grade 4 adenocarcinoma, and the parenchyma showed calcified granuloma. Patient history included benign hypertension and chronic obstructive pulmonary disease. Family history included type II diabetes and acute myocardial infarction.
72	UTRSNOT06	Library was constructed using RNA isolated from myometrial tissue removed from a 50-year-old Caucasian female during a vaginal hysterectomy. Pathology indicated residual atypical complex endometrial hyperplasia. Pathology for the associated tissue removed during dilation and curettage indicated fragments of atypical complex hyperplasia and a single microscopic focus suspicious for grade 1 adenocarcinoma. Patient history included benign breast neoplasm, hypothyroid disease, polypectomy, and arthralgia. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, hyperlipidemia, and chronic hepatitis.
73	PROSTUT08	Library was constructed using RNA isolated from prostate tumor tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 3+4). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst, and hematuria. Family history included tuberculosis, cerebrovascular disease, and arteriosclerotic coronary artery disease.
74	THYMNOT03	Library was constructed using RNA isolated from thymus tissue removed from a 21-year-old Caucasian male during a thymectomy. Pathology indicated an unremarkable thymus and a benign parathyroid adenoma in the right inferior parathyroid. Patient history included atopic dermatitis, a benign neoplasm of the parathyroid, and tobacco use. Family history included atherosclerotic coronary artery disease and benign hypertension.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
75	PENCNOT01	Library was constructed using RNA isolated from penis corpus cavernosum tissue removed from a 53-year-old male. Patient history included untreated penile carcinoma.
76	BRAUNOT01	Library was constructed using RNA isolated from caudate/putamen/nucleus accumbens tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
77	HUVELPB01	This library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells stimulated with cytokine/LPS. RNA was isolated from two pools of HUV-EC-C cells that had been treated with either 4 units/ml TNF-alpha and 2 units/ml gamma IFN for 96 hours, or 1 unit/ml IL-1 beta and 100 ng/ml LPS for 5 hours.
78	HUVENOB01	This library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells.
79	HNT2RAT01	This library was constructed at Stratagene (STR937231), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
80	BRAINOT04	This library was constructed using RNA isolated from the brain tissue of a 44-year-old Caucasian male with a cerebral hemorrhage. The tissue, which contained coagulated blood, came from the choroid plexus of the right anterior temporal lobe. Family history included coronary artery disease and myocardial infarction.
81	BRAITUT08	This library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian male during excision of cerebral meningeal tissue. Pathology indicated grade 4 fibrillary astrocytoma with focal tumoral radionecrosis. Patient history included cerebrovascular disease, deficiency anemia, hyperlipidemia and epilepsy. Family history included cerebrovascular disease and a malignant prostate neoplasm.
82	PROSNON01	This library was constructed from 4.4 million independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
83	PANCTUT01	This library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
84	BRAITUT13	This library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 68-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a meningioma in the left frontal lobe.
85	STOMFET01	This library was constructed using RNA isolated from the stomach tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
86	PROSNOT16	This library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
87	SINTNOT13	This library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
88	SINTNOT13	This library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
89	LUNGFET03	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
90	SKINBIT01	This library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
91	LUNGTUT03	This library was constructed using RNA isolated from lung tumor tissue removed from the left lower lobe of a 69-year-old Caucasian male during segmental lung resection. Pathology indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, malignant skin neoplasm, and tobacco use.
92	OVARTUT01	This library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
93	LUNGFET05	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation from anencephalus.
94	ENDANOT01	This library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
95	ESOGTUT02	This library was constructed using RNA isolated from esophageal tumor tissue obtained from a 61-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, pyloromyotomy, and regional lymph node excision. Pathology indicated an invasive grade 3 adenocarcinoma in the esophagus. Family history included atherosclerotic coronary artery disease, type II diabetes, chronic liver disease, primary cardiomyopathy, benign hypertension, and cerebrovascular disease.
96	SINIUCT01	This library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male during a total intra-abdominal colectomy and endoscopic jejunostomy. Previous surgeries included polypectomy, colonoscopy, and spinal canal exploration. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
97	NPOLNOT01	This library was constructed using RNA isolated from nasal polyp tissue removed from a 78-year-old Caucasian male during a nasal polypectomy. Pathology indicated a nasal polyp and striking eosinophilia. Patient history included asthma and nasal polyps.
98	ADRENOT09	This library was constructed using RNA isolated from left adrenal gland tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma mass in the posterior lower pole of the left kidney with invasion into the renal pelvis.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
99	BRAIUNT01	This library was constructed using RNA isolated from SK-N-MC, a neuroepithelioma cell line (ATCC HTB-10) derived from a 14-year-old Caucasian female with neuroepithelioma, with metastasis to the supra-orbital area.
100	LUNGNON03	This library was constructed from 2.56 x 1e6 independent clones from a lung tissue library. RNA was made from lung tissue removed from the left lobe a 58-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Patient also received radiation therapy to the retroperitoneum. Family history included prostate cancer, breast cancer, and acute leukemia. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228; Swaroop et al., NAR (1991) 19:1954; and Bonaldo et al., Genome Research (1996) 6:791.
101	BRADDIT02	This library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, and emphysema.
102	PLACNOT05	This library was constructed using RNA isolated from placental tissue removed from a Caucasian male fetus, who died after 18 weeks' gestation from fetal demise.
103	HELATXT03	This library was constructed using RNA isolated from a treated HeLa cell line, derived from cervical adenocarcinoma removed from a 31-year-old Black female. The cells were treated with 1 microM PMA and 100 microM cycloheximide for 24 hours.
104	COLHTUT01	This library was constructed using RNA isolated from colon tumor tissue removed from the hepatic flexure of a 55- year-old Caucasian male during right hemicolectomy, incidental appendectomy, and permanent colostomy. Pathology indicated invasive grade 3 adenocarcinoma. Patient history included benign hypertension, anxiety, abnormal blood chemistry, blepharitis, heart block, osteoporosis, acne, and hyperplasia of prostate. Family history included prostate cancer, acute myocardial infarction, stroke, and atherosclerotic coronary artery disease.
105	PLACFER01	This library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Serology was positive for CMV antibody.
106	293TF2T01	This library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
107	BRAENOT02	This library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male.
108	FTUBTUT02	This library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid and serous adenocarcinoma confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma were present. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID

NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

18. A method for treating a disease or condition associated with decreased expression of

functional CCYPR, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional CCYPR, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
HILLMAN, Jennifer L.
LAL, Preeti
TANG, Y. Tom
YUE, Henry
AU-YOUNG, Janice
BANDMAN, Olga
AZIMZAI, Yalda
YANG, Junming
LU, Dyung Aina M.
BAUGHN, Mariah R.
PATTERSON, Chandra
SHAH, Purvi

<120> CELL CYCLE AND PROLIFERATION PROTEINS

<130> PF-0722 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/145,075; 60/153,129; 60/164,647
<151> 1999-07-21; 1999-09-08; 1999-11-10

<160> 108
<170> PERL Program

<210> 1
<211> 145
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 116462CD1

<400> 1
Met Asn Gly Arg Val Asp Tyr Leu Val Thr Glu Glu Glu Ile Asn
1 5 10 15
Leu Thr Arg Gly Pro Ser Gly Leu Gly Phe Asn Ile Val Gly Gly
20 25 30
Thr Asp Gln Gln Tyr Val Ser Asn Asp Ser Gly Ile Tyr Val Ser
35 40 45
Arg Ile Lys Glu Asn Gly Ala Ala Ala Leu Asp Gly Arg Leu Gln
50 55 60
Glu Gly Asp Lys Ile Leu Ser Val Asn Gly Gln Asp Leu Lys Asn
65 70 75
Leu Leu His Gln Asp Ala Val Asp Leu Phe Arg Asn Ala Gly Tyr
80 85 90
Ala Val Ser Leu Arg Val Gln His Arg Leu Gln Val Gln Asn Gly
95 100 105
Pro Ile Gly His Arg Gly Glu Gly Asp Pro Ser Gly Ile Pro Ile
110 115 120
Phe Met Val Leu Val Pro Val Phe Ala Leu Thr Met Val Ala Ala
125 130 135
Trp Ala Phe Met Arg Tyr Arg Gln Gln Leu
140 145

<210> 2
<211> 340
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1210462CD1

<400> 2
Met Leu Thr Gln Leu Lys Ala Lys Ser Glu Gly Lys Leu Ala Lys
1 5 10 15
Gln Ile Cys Lys Val Val Leu Asp His Phe Glu Lys Gln Tyr Ser
20 25 30
Lys Glu Leu Gly Asp Ala Trp Asn Thr Val Arg Glu Ile Leu Thr
35 40 45
Ser Pro Ser Cys Trp Gln Tyr Ala Val Leu Leu Asn Arg Phe Asn
50 55 60
Tyr Pro Phe Glu Leu Glu Lys Asp Leu His Leu Lys Gly Tyr His
65 70 75
Thr Leu Ser Gln Gly Ser Leu Pro Asn Tyr Pro Lys Ser Val Lys
80 85 90
Cys Tyr Leu Ser Arg Thr Pro Gly Arg Ile Pro Ser Glu Arg His
95 100 105
Gln Ile Gly Asn Leu Lys Lys Tyr Tyr Leu Leu Asn Ala Ala Ser
110 115 120
Leu Leu Pro Val Leu Ala Leu Glu Leu Arg Asp Gly Glu Lys Val
125 130 135
Leu Asp Leu Cys Ala Ala Pro Gly Gly Lys Ser Ile Ala Leu Leu
140 145 150
Gln Cys Ala Cys Pro Gly Tyr Leu His Cys Asn Glu Tyr Asp Ser
155 160 165
Leu Arg Leu Arg Trp Leu Arg Gln Thr Leu Glu Ser Phe Ile Pro
170 175 180
Gln Pro Leu Ile Asn Val Ile Lys Val Ser Glu Leu Asp Gly Arg
185 190 195
Lys Met Gly Asp Ala Gln Pro Glu Met Phe Asp Lys Val Leu Val
200 205 210
Asp Ala Pro Cys Ser Asn Asp Arg Ser Trp Leu Phe Ser Ser Asp
215 220 225
Ser Gln Lys Ala Ser Cys Arg Ile Ser Gln Arg Arg Asn Leu Pro
230 235 240
Leu Leu Gln Ile Glu Leu Leu Arg Ser Ala Ile Lys Ala Leu Arg
245 250 255
Pro Gly Gly Ile Leu Val Tyr Ser Thr Cys Thr Leu Ser Lys Ala
260 265 270
Glu Asn Gln Asp Val Ile Ser Glu Ile Leu Asn Ser His Gly Asn
275 280 285
Ile Met Pro Met Asp Ile Lys Gly Ile Ala Arg Thr Cys Ser His
290 295 300
Asp Phe Thr Phe Ala Pro Thr Gly Gln Glu Cys Gly Leu Leu Val
305 310 315
Ile Pro Asp Lys Gly Lys Ala Trp Gly Pro Met Tyr Val Ala Lys
320 325 330
Leu Lys Lys Ser Trp Ser Thr Gly Lys Trp
335 340

<210> 3
<211> 418
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1305252CD1

<400> 3

```
Met Leu Tyr Leu Glu Asp Tyr Leu Glu Met Ile Glu Gln Leu Pro
 1               5                  10                  15
Met Asp Leu Arg Asp Arg Phe Thr Glu Met Arg Glu Met Asp Leu
            20                  25                  30
Gln Val Gln Asn Ala Met Asp Gln Leu Glu Gln Arg Val Ser Glu
            35                  40                  45
Phe Phe Met Asn Ala Lys Lys Asn Lys Pro Glu Trp Arg Glu Glu
            50                  55                  60
Gln Met Ala Ser Ile Lys Lys Asp Tyr Tyr Lys Ala Leu Glu Asp
            65                  70                  75
Ala Asp Glu Lys Val Gln Leu Ala Asn Gln Ile Tyr Asp Leu Val
            80                  85                  90
Asp Arg His Leu Arg Lys Leu Asp Gln Glu Leu Ala Lys Phe Lys
            95                 100                 105
Met Glu Leu Glu Ala Asp Asn Ala Gly Ile Thr Glu Ile Leu Glu
           110                 115                 120
Arg Arg Ser Leu Glu Leu Asp Thr Pro Ser Gln Pro Val Asn Asn
           125                 130                 135
His His Ala His Ser His Thr Pro Val Glu Lys Arg Lys Tyr Asn
           140                 145                 150
Pro Thr Ser His His Thr Thr Thr Asp His Ile Pro Glu Lys Lys
           155                 160                 165
Phe Lys Ser Glu Ala Leu Leu Ser Thr Leu Thr Ser Asp Ala Ser
           170                 175                 180
Lys Glu Asn Thr Leu Gly Cys Arg Asn Asn Asn Ser Thr Ala Ser
           185                 190                 195
Ser Asn Asn Ala Tyr Asn Val Asn Ser Ser Gln Pro Leu Gly Ser
           200                 205                 210
Tyr Asn Ile Gly Ser Leu Ser Ser Gly Thr Gly Ala Gly Ala Ile
           215                 220                 225
Thr Met Ala Ala Ala Gln Ala Val Gln Ala Thr Ala Gln Met Lys
           230                 235                 240
Glu Gly Arg Arg Thr Ser Ser Leu Lys Ala Ser Tyr Glu Ala Phe
           245                 250                 255
Lys Asn Asn Asp Phe Gln Leu Gly Lys Glu Phe Ser Met Ala Arg
           260                 265                 270
Glu Thr Val Gly Tyr Ser Ser Ser Ser Ala Leu Met Thr Thr Leu
           275                 280                 285
Thr Gln Asn Ala Ser Ser Ser Ala Ala Asp Ser Arg Ser Gly Arg
           290                 295                 300
Lys Ser Lys Asn Asn Asn Lys Ser Ser Ser Gln Gln Ser Ser Ser
           305                 310                 315
Ser Ser Ser Ser Ser Ser Leu Ser Ser Cys Ser Ser Ser Ser Thr
           320                 325                 330
Val Val Gln Glu Ile Ser Gln Gln Thr Thr Val Val Pro Glu Ser
           335                 340                 345
Asp Ser Asn Ser Gln Val Asp Trp Thr Tyr Asp Pro Asn Glu Pro
           350                 355                 360
Arg Tyr Cys Ile Cys Asn Gln Val Ser Tyr Gly Glu Met Val Gly
           365                 370                 375
Cys Asp Asn Gln Asp Cys Pro Ile Glu Trp Phe His Tyr Gly Cys
           380                 385                 390
Val Gly Leu Thr Glu Ala Pro Lys Gly Lys Trp Tyr Cys Pro Gln
           395                 400                 405
Cys Thr Ala Ala Met Lys Arg Arg Gly Ser Arg His Lys
           410                 415
```

<210> 4
<211> 297
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 1416289CD1

<400> 4
Met Ala Tyr Asn Val Ile Ile Ile Tyr Phe Asn Phe Arg Cys Leu
1 5 10 15
Glu Trp Leu Leu Asn Asn Leu Met Thr His Gln Asn Val Glu Leu
20 25 30
Phe Lys Glu Leu Ser Ile Asn Val Met Lys Gln Leu Ile Gly Ser
35 40 45
Ser Asn Leu Phe Val Met Gln Val Glu Met Asp Ile Tyr Thr Ala
50 55 60
Leu Lys Lys Trp Met Phe Leu Gln Leu Val Pro Ser Trp Asn Gly
65 70 75
Ser Leu Lys Gln Leu Leu Thr Glu Thr Asp Val Trp Phe Ser Lys
80 85 90
Gln Arg Lys Asp Phe Glu Gly Met Ala Phe Leu Glu Thr Glu Gln
95 100 105
Gly Lys Pro Phe Val Ser Val Phe Arg His Leu Arg Leu Gln Tyr
110 115 120
Ile Ile Ser Asp Leu Ala Ser Ala Arg Ile Ile Glu Gln Asp Ala
125 130 135
Val Val Pro Ser Glu Trp Leu Ser Ser Val Tyr Lys Gln Gln Trp
140 145 150
Phe Ala Met Leu Arg Ala Glu Gln Asp Ser Glu Val Gly Pro Gln
155 160 165
Glu Ile Asn Lys Glu Glu Leu Glu Gly Asn Ser Met Arg Cys Gly
170 175 180
Arg Lys Leu Ala Lys Asp Gly Glu Tyr Cys Trp Arg Trp Thr Gly
185 190 195
Phe Asn Phe Gly Phe Asp Leu Leu Val Thr Tyr Thr Asn Arg Tyr
200 205 210
Ile Ile Phe Lys Arg Asn Thr Leu Asn Gln Pro Cys Ser Gly Ser
215 220 225
Val Ser Leu Gln Pro Arg Arg Ser Ile Ala Phe Arg Leu Arg Leu
230 235 240
Ala Ser Phe Asp Ser Ser Gly Lys Leu Ile Cys Ser Arg Thr Thr
245 250 255
Gly Tyr Gln Ile Leu Thr Leu Glu Lys Asp Gln Glu Gln Val Val
260 265 270
Met Asn Leu Asp Ser Arg Leu Leu Ile Phe Pro Leu Tyr Ile Cys
275 280 285
Cys Asn Phe Leu Tyr Ile Ser Pro Glu Lys Lys Asn
290 295

<210> 5
<211> 184
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1558289CD1

<400> 5
Met Glu Ser Phe Ser Ser Lys Ser Leu Ala Leu Gln Ala Glu Lys
1 5 10 15
Lys Leu Leu Ser Lys Met Ala Gly Arg Ser Val Ala His Leu Phe
20 25 30
Ile Asp Glu Thr Ser Ser Glu Val Leu Asp Glu Leu Tyr Arg Val
35 40 45
Ser Lys Glu Tyr Thr His Ser Arg Pro Gln Ala Gln Arg Val Ile
50 55 60
Lys Asp Leu Ile Lys Val Ala Ile Lys Val Ala Val Leu His Arg
65 70 75

```
Asn Gly Ser Phe Gly Pro Ser Glu Leu Ala Leu Ala Thr Arg Phe
                80                  85                  90
Arg Gln Lys Leu Arg Gln Gly Ala Met Thr Ala Leu Ser Phe Gly
                95                 100                 105
Glu Val Asp Phe Thr Phe Glu Ala Ala Val Leu Ala Gly Leu Leu
               110                 115                 120
Thr Glu Cys Arg Asp Val Leu Leu Glu Leu Val Glu His His Leu
               125                 130                 135
Thr Pro Lys Ser His Gly Arg Ile Arg His Val Phe Asp His Phe
               140                 145                 150
Ser Asp Pro Gly Leu Leu Thr Ala Leu Tyr Gly Pro Asp Phe Thr
               155                 160                 165
Gln His Leu Gly Lys Ile Cys Asp Gly Leu Arg Lys Leu Leu Asp
               170                 175                 180
Glu Gly Lys Leu
```

<210> 6
<211> 173
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1577739CD1

```
<400> 6
Met Asp Val Arg Arg Val Leu Val Lys Ala Glu Met Glu Lys Phe
  1                 5                  10                  15
Leu Gln Asn Lys Glu Leu Phe Ser Ser Leu Lys Lys Gly Lys Ile
                20                  25                  30
Cys Cys Cys Cys Arg Ala Lys Phe Pro Leu Phe Ser Trp Pro Pro
                35                  40                  45
Ser Cys Leu Phe Cys Lys Arg Ala Val Cys Thr Ser Cys Ser Ile
                50                  55                  60
Lys Met Lys Met Pro Ser Lys Lys Phe Gly His Ile Pro Val Tyr
                65                  70                  75
Thr Leu Gly Phe Glu Ser Pro Gln Arg Val Ser Ala Ala Lys Thr
                80                  85                  90
Ala Pro Ile Gln Arg Arg Asp Ile Phe Gln Ser Leu Gln Gly Pro
                95                 100                 105
Gln Trp Gln Ser Val Glu Glu Ala Phe Pro His Ile Tyr Ser His
               110                 115                 120
Gly Cys Val Leu Lys Asp Val Cys Ser Glu Cys Thr Ser Phe Val
               125                 130                 135
Ala Asp Val Val Arg Ser Ser Arg Lys Ser Val Asp Val Leu Asn
               140                 145                 150
Thr Thr Pro Arg Arg Ser Arg Gln Thr Gln Ser Leu Tyr Ile Pro
               155                 160                 165
Asn Thr Arg Thr Leu Asp Phe Lys
               170
```

<210> 7
<211> 591
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1752768CD1

```
<400> 7
Met Val Pro Val Ala Val Thr Ala Ala Val Ala Pro Val Leu Ser
  1                 5                  10                  15
Ile Asn Ser Asp Phe Ser Asp Leu Arg Glu Ile Lys Lys Gln Leu
```

```
                20                  25                  30
Leu Leu Ile Ala Gly Leu Thr Arg Glu Arg Gly Leu Leu His Ser
                35                  40                  45
Ser Lys Trp Ser Ala Glu Leu Ala Phe Ser Leu Pro Ala Leu Pro
                50                  55                  60
Leu Ala Glu Leu Gln Pro Pro Pro Pro Ile Thr Glu Glu Asp Ala
                65                  70                  75
Gln Asp Met Asp Ala Tyr Thr Leu Ala Lys Ala Tyr Phe Asp Val
                80                  85                  90
Lys Glu Tyr Asp Arg Ala Ala His Phe Leu His Gly Cys Asn Ser
                95                  100                 105
Lys Lys Ala Tyr Phe Leu Tyr Met Tyr Ser Arg Tyr Leu Ser Gly
                110                 115                 120
Glu Lys Lys Lys Asp Asp Glu Thr Val Asp Ser Leu Gly Pro Leu
                125                 130                 135
Glu Lys Gly Gln Val Lys Asn Glu Ala Leu Arg Glu Leu Arg Val
                140                 145                 150
Glu Leu Ser Lys Lys His Gln Ala Arg Glu Leu Asp Gly Phe Gly
                155                 160                 165
Leu Tyr Leu Tyr Gly Val Val Leu Arg Lys Leu Asp Leu Val Lys
                170                 175                 180
Glu Ala Ile Asp Val Phe Val Glu Ala Thr His Val Leu Pro Leu
                185                 190                 195
His Trp Gly Ala Trp Leu Glu Leu Cys Asn Leu Ile Thr Asp Lys
                200                 205                 210
Glu Met Leu Lys Phe Leu Ser Leu Pro Asp Thr Trp Met Lys Glu
                215                 220                 225
Phe Phe Leu Ala His Ile Tyr Thr Glu Leu Gln Leu Ile Glu Glu
                230                 235                 240
Ala Leu Gln Lys Tyr Gln Asn Leu Ile Asp Val Gly Phe Ser Lys
                245                 250                 255
Ser Ser Tyr Ile Val Ser Gln Ile Ala Val Ala Tyr His Asn Ile
                260                 265                 270
Arg Asp Ile Asp Lys Ala Leu Ser Ile Phe Asn Glu Leu Arg Lys
                275                 280                 285
Gln Asp Pro Tyr Arg Ile Glu Asn Met Asp Thr Phe Ser Asn Leu
                290                 295                 300
Leu Tyr Val Arg Ser Met Lys Ser Glu Leu Ser Tyr Leu Ala His
                305                 310                 315
Asn Leu Cys Glu Ile Asp Lys Tyr Arg Val Glu Thr Cys Cys Val
                320                 325                 330
Ile Gly Asn Tyr Tyr Ser Leu Arg Ser Gln His Glu Lys Ala Ala
                335                 340                 345
Leu Tyr Phe Gln Arg Ala Leu Lys Leu Asn Pro Arg Tyr Leu Gly
                350                 355                 360
Ala Trp Thr Leu Met Gly His Glu Tyr Met Glu Met Lys Asn Thr
                365                 370                 375
Ser Ala Ala Ile Gln Ala Tyr Arg His Ala Ile Glu Val Asn Lys
                380                 385                 390
Arg Asp Tyr Arg Ala Trp Tyr Gly Leu Gly Gln Thr Tyr Glu Ile
                395                 400                 405
Leu Lys Met Pro Phe Tyr Cys Leu Tyr Tyr Cys Arg Arg Ala His
                410                 415                 420
Gln Leu Arg Pro Asn Asp Ser Arg Met Leu Val Ala Leu Gly Glu
                425                 430                 435
Cys Tyr Glu Lys Leu Asn Gln Leu Val Glu Ala Lys Lys Cys Tyr
                440                 445                 450
Trp Arg Ala Tyr Ala Val Gly Asp Val Glu Lys Met Ala Leu Val
                455                 460                 465
Lys Leu Ala Lys Leu His Glu Gln Leu Thr Glu Ser Glu Gln Ala
                470                 475                 480
Ala Gln Cys Tyr Ile Lys Tyr Ile Gln Asp Ile Tyr Ser Cys Gly
                485                 490                 495
```

Glu Ile Val Glu His Leu Glu Glu Ser Thr Ala Phe Arg Tyr Leu
500 505 510
Ala Gln Tyr Tyr Phe Lys Cys Lys Leu Trp Asp Glu Ala Ser Thr
515 520 525
Cys Ala Gln Lys Cys Cys Ala Phe Asn Asp Thr Arg Glu Glu Gly
530 535 540
Lys Ala Leu Leu Arg Gln Ile Leu Gln Leu Arg Asn Gln Gly Glu
545 550 555
Thr Pro Thr Thr Glu Val Pro Ala Pro Phe Phe Leu Pro Ala Ser
560 565 570
Leu Ser Ala Asn Asn Thr Pro Thr Arg Arg Val Ser Pro Leu Asn
575 580 585
Leu Ser Ser Val Thr Pro
590

<210> 8
<211> 463
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1887228CD1

<400> 8
Met Pro Leu Leu Asn Trp Val Ala Leu Lys Pro Ser Gln Ile Thr
1 5 10 15
Gly Thr Val Phe Thr Glu Leu Asn Asp Glu Lys Val Leu Gln Glu
20 25 30
Leu Asp Met Ser Asp Phe Glu Glu Gln Phe Lys Thr Lys Ser Gln
35 40 45
Gly Pro Ser Leu Asp Leu Ser Ala Leu Lys Ser Lys Ala Ala Gln
50 55 60
Lys Ala Pro Ser Lys Ala Thr Leu Ile Glu Ala Asn Arg Ala Lys
65 70 75
Asn Leu Ala Ile Thr Leu Arg Lys Gly Asn Leu Gly Ala Glu Arg
80 85 90
Ile Cys Gln Ala Ile Glu Ala Tyr Asp Leu Gln Ala Leu Gly Leu
95 100 105
Asp Phe Leu Glu Leu Leu Met Arg Phe Leu Pro Thr Glu Tyr Glu
110 115 120
Arg Ser Leu Ile Thr Arg Phe Glu Arg Glu Gln Arg Pro Met Glu
125 130 135
Glu Leu Ser Glu Glu Asp Arg Phe Met Leu Cys Phe Ser Arg Ile
140 145 150
Pro Arg Leu Pro Glu Arg Met Thr Thr Leu Thr Phe Leu Gly Asn
155 160 165
Phe Pro Asp Thr Ala Gln Leu Leu Met Pro Gln Leu Asn Ala Ile
170 175 180
Ile Ala Ala Ser Met Ser Ile Lys Ser Ser Asp Lys Leu Arg Gln
185 190 195
Ile Leu Glu Ile Val Leu Ala Phe Gly Asn Tyr Met Asn Ser Ser
200 205 210
Lys Arg Gly Ala Ala Tyr Gly Phe Arg Leu Gln Ser Leu Asp Ala
215 220 225
Leu Leu Glu Met Lys Ser Thr Asp Arg Lys Gln Thr Leu Leu His
230 235 240
Tyr Leu Val Lys Val Ile Ala Glu Lys Tyr Pro Gln Leu Thr Gly
245 250 255
Phe His Ser Asp Leu His Phe Leu Asp Lys Ala Gly Ser Val Ser
260 265 270
Leu Asp Ser Val Leu Ala Asp Val Arg Ser Leu Gln Arg Gly Leu
275 280 285
Glu Leu Thr Gln Arg Glu Phe Val Arg Gln Asp Asp Cys Met Val

290 295 300
Leu Lys Glu Phe Leu Arg Ala Asn Ser Pro Thr Met Asp Lys Leu
305 310 315
Leu Ala Asp Ser Lys Thr Ala Gln Glu Ala Phe Glu Ser Val Val
320 325 330
Glu Tyr Phe Gly Glu Asn Pro Lys Thr Thr Ser Pro Gly Leu Phe
335 340 345
Phe Ser Leu Phe Ser Arg Phe Ile Lys Ala Tyr Lys Lys Ala Glu
350 355 360
Gln Glu Val Glu Gln Trp Lys Lys Glu Ala Ala Ala Gln Glu Ala
365 370 375
Gly Ala Asp Thr Pro Gly Lys Gly Glu Pro Pro Ala Pro Lys Ser
380 385 390
Pro Pro Lys Ala Arg Arg Pro Gln Met Asp Leu Ile Ser Glu Leu
395 400 405
Lys Arg Arg Gln Gln Lys Glu Pro Leu Ile Tyr Glu Ser Asp Arg
410 415 420
Asp Gly Ala Ile Glu Asp Ile Ile Thr Asp Leu Arg Asn Gln Pro
425 430 435
Tyr Ile Arg Ala Asp Thr Gly Arg Arg Ser Ala Arg Arg Arg Pro
440 445 450
Pro Gly Pro Pro Leu Gln Val Thr Ser Asp Leu Ser Leu
455 460

<210> 9
<211> 270
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1988468CD1

<400> 9
Met Ala Asp His Met Met Ala Met Asn His Gly Arg Phe Pro Asp
1 5 10 15
Gly Thr Asn Gly Leu His His His Pro Ala His Arg Met Gly Met
20 25 30
Gly Gln Phe Pro Ser Pro His His His Gln Gln Gln Gln Pro Gln
35 40 45
His Ala Phe Asn Ala Leu Met Gly Glu His Ile His Tyr Gly Ala
50 55 60
Gly Asn Met Asn Ala Thr Ser Gly Ile Arg His Ala Met Gly Pro
65 70 75
Gly Thr Val Asn Gly Gly His Pro Pro Ser Ala Leu Ala Pro Ala
80 85 90
Ala Arg Phe Asn Asn Ser Gln Phe Met Gly Pro Pro Val Ala Ser
95 100 105
Gln Gly Gly Ser Leu Pro Ala Ser Met Gln Leu Gln Lys Leu Asn
110 115 120
Asn Gln Tyr Phe Asn His His Pro Tyr Pro His Asn His Tyr Met
125 130 135
Pro Asp Leu His Pro Ala Ala Gly His Gln Met Asn Gly Thr Asn
140 145 150
Gln His Phe Arg Asp Cys Asn Pro Lys His Ser Gly Gly Ser Ser
155 160 165
Thr Pro Gly Gly Ser Gly Gly Ser Ser Thr Pro Gly Gly Ser Gly
170 175 180
Ser Ser Ser Gly Gly Gly Ala Gly Ser Ser Asn Ser Gly Gly Gly
185 190 195
Ser Gly Ser Gly Asn Met Pro Ala Ser Val Ala His Val Pro Ala
200 205 210
Ala Met Leu Pro Pro Asn Val Ile Asp Thr Asp Phe Ile Asp Glu
215 220 225